METHODS OF REFOLDING MAMMALIAN GLYCOSYLTRANSFERASES

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 60/542,210, filed February 4, 2004, and of U.S. Provisional Application No. 60/599,406, filed August 6, 2004, and of U.S. Provisional Application No. 60/627,406, filed November 12, 2004; each of which are herein incorporated by reference for all purposes.

FIELD OF INVENTION

[0002] The present invention provides methods of refolding mammalian glycosyltransferases that have been produced in bacterial cells, including glycosyltransferase mutants that have enhanced ability to be refolded, and methods to use such refolded glycosyltransferases. The invention also provides methods of refolding more than one glycosyltransferase in a single vessel, methods to use such refolded glycosyltransferases, and reaction mixtures comprising the refolded glycosyltransferases.

BACKGROUND OF THE INVENTION

[0003] Eukaryotic organisms synthesize oligosaccharide structures or glycoconjugates, such as glycolipids or glycoproteins, that are commercially and therapeutically useful. *In vitro* synthesis of oligosaccharides or glycoconjugates can be carried out using recombinant eukaryotic glycosyltransferases. The most efficient method to produce recombinant eukaryotic glycosyltransferases for oligosaccharide synthesis is to express the protein in bacteria. However, in bacteria, many eukaryotic glycosyltransferases are expressed as insoluble proteins in bacterial inclusion bodies, and yields of active protein from the inclusion bodies can be very low. Thus, there is a need for improved methods to produce eukaryotic glycosyltransferases in bacteria. The present invention solves this and other needs.

BRIEF SUMMARY OF THE INVENTION

[0004] The present invention provides a method of refolding an insoluble, recombinant, eukaryotic glycosyltransferase that comprises a maltose binding protein domain (MBD). The insoluble, recombinant, eukaryotic glycosyltransferase is solubilized in a solubilization buffer and then contacted with a refolding buffer comprising a redox couple, such that the refolded eukaryotic glycosyltransferase catalyzes the transfer of a sugar from a donor substrate to an acceptor substrate. In one embodiment, the eukaryotic glycosyltransferase is truncated to

remove all or a portion of a stem region of the protein. In another embodiment, an unpaired cysteine in the eukaryotic glycosyltransferase is removed by substitution with a non-cysteine amino acid. In a further embodiment an unpaired cysteine in the eukaryotic glycosyltransferase is removed by substitution with a non-cysteine amino acid and the the eukaryotic glycosyltransferase is also truncated to remove all or a portion of a stem region of the protein.

[0005] In one embodiment, the eukaryotic glycosyltransferase is selected from the group consisting of GnT1, GalT1, StIII Gal3, St3GalI, St6 GalNAcTI, Core GalITI, GalNAcT2.

[0006] In one embodiment, the eukaryotic glycosyltransferase further comprises a purification domain, for example, a starch binding domain (SBD), a thioredoxin domain, a SUMO domain, a poly-His domain, a myc epitope domain, and a glutathione-S-transferase domain.

[0007] In one embodiment, the eukaryotic glycosyltransferase further comprises a self cleaving domain.

[0008] In one embodiment, the eukaryotic glycosyltransferase is expressed in a bacterial host cell as an insoluble inclusion body.

[0009] In one embodiment, a second insoluble, recombinant eukaryotic glycosyltransferase is refolded with the first eukaryotic glycosyltransferase. In a further embodiment, a third insoluble, recombinant eukaryotic glycosyltransferase is refolded with the first eukaryotic glycosyltransferase and the second eukaryotic glycosyltransferase. Additional insoluble, recombinant eukaryotic glycosyltransferase can be added and refolded together depending on the needs of the user, *e.g.*, refolding of 4, 5, 6, 7, 8, 9, or 10 glycosyltransferases together.

[0010] In one embodiment, the redox couple is selected from the group consisting of reduced glutathione/oxidized glutathione (GSH/GSSG) and cysteine/ cystamine.

[0011] In one embodiment, the acceptor substrate is selected from a protein, a peptide, a glycoprotein, and a glycopeptide.

[0012] In one embodiment, the eukaryotic glycosyltransferase is a sialyltransferase. Sialyltransferases can include, *e.g.*, StIII Gal3, St3GalI, St6 GalNAcTI. In a further aspect, the donor substrate is a CMP-sialic acid PEG molecule and the acceptor substrate is selected from a protein, a peptide, a glycoprotein, and a glycopeptide.

[0013] The present invention also provides a recombinant, eukaryotic glycosyltransferase, in which a stem anchor region and a transmembrane domain have been deleted from the protein, and wherein the glycosyltransferase is fused in frame to a maltose binding protein (MBP) domain. In one embodiment, the fusion of the recombinant, eukaryotic glycosyltransferase and the MBP domain is expressed as an insoluble inclusion body in bacteria, *e.g.*, *E. coli.*

[0014] In one embodiment, all or a portion of the stem region is deleted from the recombinant, eukaryotic glycosyltransferase. In another embodiment, an unpaired cysteine in the recombinant, eukaryotic glycosyltransferase is removed by substitution with a non-cysteine amino acid.

[0015] In a further embodiment, the recombinant, eukaryotic glycosyltransferase is one of the following: a GnT1 protein, a GalT1 protein, an StIII Gal3 protein, an St3GalI protein, an St6 GalNAcTI protein, a Core GalITI protein, or a GalNAcT2 protein.

[0016] In one embodiment, the recombinant, eukaryotic glycosyltransferase is a GnT1 protein. In one aspect, the GnT1 protein is a truncated human GnT1 protein selected from GnT1 Δ35 and GnT1Δ103. In another aspect, the GnT1 protein is a human GnT1 protein comprising an unpaired cysteine substitution selected from the group consisting of CYS121ALA, CYS121ASP, and ARG120ALA,CYS121HIS. In a further aspect, the GnT1 protein is both truncated and has had an unpaired cysteine residue removed by a substitution mutation.

[0017] In one embodiment, the recombinant, eukaryotic glycosyltransferase is a GalT1 protein. In one aspect, the GalT1 protein is a truncated bovine GalT1 protein selected from GalT1 Δ70 and GalT1 Δ129. In another aspect, the GalT1 protein is a bovine GalT1 protein comprising an unpaired cysteine substitution of CYS342THR. In a further aspect, the GalT1 protein is both truncated and has had an unpaired cysteine residue removed by a substitution mutation.

[0018] In one embodiment, the recombinant, eukaryotic glycosyltransferase is an ST3GalIII protein. In one aspect, the ST3GalIII protein is a truncated rat ST3GalIII protein selected from ST3GalIII Δ28, ST3GalIII Δ73, ST3GalIII Δ85 and ST3GalIII Δ86. In another aspect, the ST3GalIII protein comprises an amino acid substitution for an unpaired cysteine residue. In a further aspect, the ST3GalIII protein is both truncated and has had an unpaired cysteine residue removed by a substitution mutation.

[0019] In one embodiment, the recombinant, eukaryotic glycosyltransferase of claim 15, wherein the glycosyltransferase is a Core1 GalT1 protein. In one aspect, the Core1 GalT1 protein is a truncated Drosophila protein or a truncated human protein. In another aspect, a Drosophila or human Core1 GalT1 protein comprises an amino acid substitution for an unpaired cysteine residue. In a further aspect, a Drosophila or human Core1 GalT1 protein is both truncated and has had an unpaired cysteine residue removed by a substitution mutation.

[0020] In one embodiment, the recombinant, eukaryotic glycosyltransferase is an ST3Gal1 protein. In one aspect, the ST3Gal1 protein is a truncated human protein selected from ST3Gal1 Δ29, ST3Gal1 Δ45, and ST3Gal1 Δ56. In another aspect, the ST3Gal1 protein comprises an amino acid substitution for an unpaired cysteine residue. In a further aspect, the ST3Gal1 protein is both truncated and has had an unpaired cysteine residue removed by a substitution mutation.

[0021] In one embodiment, the recombinant, eukaryotic glycosyltransferase is an ST6GalNAc1 protein. In one aspect, the ST6GalNAc1 protein is a truncated mouse protein, a truncated chicken protein or a truncated human protein, *e.g.* one of the truncation listed in Table 14. In another aspect, a mouse, chicken, or human ST6GalNAc1 protein comprises an amino acid substitution for an unpaired cysteine residue. In a further aspect, a mouse, chicken, or human ST6GalNAc1 protein is both truncated and has had an unpaired cysteine residue removed by a substitution mutation.

[0022] In one embodiment, the recombinant, eukaryotic glycosyltransferase of claim 15, wherein the glycosyltransferase is an GalNAcT2 protein. In one aspect, the GalNAcT2 protein is a truncated human protein selected from GalNAcT2 Δ40, GalNAcT2 Δ51, GalNAcT2 Δ74 and GalNAcT2 Δ95. In another aspect, the human GalNAcT2 protein comprises an amino acid substitution for an unpaired cysteine residue. In a further aspect, the human GalNAcT2 protein is both truncated and has had an unpaired cysteine residue removed by a substitution mutation.

[0023] The present invention also provides a method of remodeling a protein, a peptide, a glycoprotein, or a glycopeptide using the recombinant, eukaryotic glycosyltransferases listed above, after the proteins have been refolded and have enzymatic activity.

[0024] The present invention provides improved methods to refold insoluble eukaryotic glycosyltransferases in an active form and also provides glycosyltransferases, *e.g.*, N-acetylglucosaminyltransferase I (GnTI) enzymes that have enhanced refolding properties.

[0025] In one aspect, the invention provides a recombinant eukaryotic N-acetylglucosaminyltransferase I (GnTI) enzyme, that has been mutated to replace an unpaired cysteine residue with an amino acid that enhances refolding of the enzyme from an insoluble precipitate, *e.g.*, bacterial inclusion bodies. The GnT1 enzyme includes at least the catalytic domain of the GnT1 enzyme. The GnT1 enzyme is biologically active, *i.e.*, able to catalyze the transfer of a donor substrate to an acceptor substrate.

[0026] In one embodiment, the GnTI enzyme is a human protein. Some mutations of the CYS121 residue in human GnT1 enhance refolding. Those mutants include *e.g.*, CYS121SER mutation, a CYS121ALA mutation, CYS121ASP mutation, and a double mutant, ARG120ALA, CYS121HIS. Representative sequences of GnT1 mutants are shown in Figures 7-11. In other eukaryotes, *e.g.*, similar mutations of an unpaired cysteine residue, CYS123, enhance refolding of the GnT1 enzyme.

[0027] In another embodiment, the GnTI enzyme also includes an amino acid tag, *e.g.*, a maltose binding protein (MBP), a polyhistidine tag, a glutathione S transferase (GST), a starch binding protein (SBP), and a myc epitope.

[0028] In another aspect, the invention provides nucleic acids encoding a recombinant eukaryotic GnTI enzyme, that has been mutated to replace an unpaired cysteine residue with an amino acid that enhances refolding of the enzyme from an insoluble precipitate, *e.g.*, bacterial inclusion bodies. As above, the encoded GnT1 enzyme includes at least the catalytic domain of the GnT1 enzyme, and is biologically active, *i.e.*, able to catalyze the transfer of a donor substrate to an acceptor substrate.

[0029] In one embodiment, the nucleic acids encode a human GnTI enzyme. Some mutations of the CYS121 residue in human GnT1 enhance refolding. Those mutants include *e.g.*, CYS121SER mutation, a CYS121ALA mutation, CYS121ASP mutation, and a double mutant, ARG120ALA, CYS121HIS. Representative nucleic acids sequences of GnT1 mutant proteins and nucleic acids are shown in Figures 7-11. In other eukaryotes, *e.g.*, similar mutations of an unpaired cysteine residue, CYS123, enhance refolding of the GnT1 enzyme.

[0030] In a further embodiment, the encoded GnTI enzyme also includes an amino acid tag, *e.g.*, a maltose binding protein (MBP), a polyhistidine tag, a glutathione S transferase (GST), a starch binding protein (SBP), and a myc epitope.

[0031] The invention also includes expression vectors that include the mutated GnT1 nucleic acids, host cells that include the GnT1 expression vectors, and methods of producing the mutated GnT1 enzymes using the host/expression vector system.

[0032] In another embodiment, the invention provides a method of adding N-acetylglucosamine residues to an acceptor molecule with a terminal mannose residue, by contacting the acceptor molecule with an activated N-acetylglucosamine molecule and a eukaryotic GnTI enzyme that has been mutated to enhance refolding. The acceptor molecule can be *e.g.*, a polysaccharide, an oligosaccharide, a glycolipid, or a glycoprotein.

[0033] In another aspect, the invention provides a method of refolding at least two insoluble, recombinant eukaryotic glycosyltransferase proteins in a single vessel, by contacting the glycosyltransferases with a refolding buffer that includes a redox couple. After refolding, at least two of the refolded glycosyltransferases have biological activity, *e.g.*, are able to catalyze the transfer of a donor substrate to an acceptor substrate.

[0034] The refolding buffer can also include a detergent, or a chaotropic agent, or arginine, or PEG. In some embodiments the pH of the refolding buffer is between 6.0 and 10.0. In one embodiment, the pH of the refolding buffer is between 6.5 and 8.0. In another embodiment, the pH of the refolding buffer is between 8.0 and 9.0.

[0035] In another embodiment, the glycosyltransferases include an amino acid tag, *e.g.*, a maltose binding protein (MBP), a polyhistidine tag, a glutathione S transferase (GST), a starch binding protein (SBP), and a myc epitope

[0036] In one embodiment, more than one glycosyltransferase from an N-linked glycan biosynthetic pathway are refolded together.

[0037] In one embodiment, a sialyltransferase is refolded with another glycosyltransferase using the methods of the invention.

[0038] In one embodiment, an N-acetylglucosaminyltransferase is refolded with another glycosyltransferase using the methods of the invention.

[0039] In one embodiment, a galactosyltransferase is refolded with another glycosyltransferase using the methods of the invention.

[0040] In another embodiment, a sialyltransferase, an N-acetylglucosaminyltransferase, and a galactosyltransferase are refolded together in a single vessel using the methods of the invention.

[0041] In one embodiment, more than one glycosyltransferase from an O-linked glycan biosynthetic pathway are refolded together. In a further embodiment, a first enzyme is an N-acetylgalactosaminyltransferase. In a preferred embodiment, a first enzyme is an N-acetylglucosaminyltransferase 2 (GalNAcT2).

[0042] The present invention also provides a reaction mixture including a recombinant eukaryotic GnTI enzyme, that has been mutated to replace an unpaired cysteine residue with an amino acid that enhances refolding of the enzyme from an insoluble precipitate, *e.g.*, bacterial inclusion bodies and at least one other glycosyltransferase that have been refolded in the same vessel. The second glycosyltransferase can be *e.g.*, a sialyltransferase or a galactosyltransferase. In one embodiment, the reaction mixture includes the mutated eukaryotic GnT1 enzyme, a sialyltransferase, and a galactosyltransferase. The reaction mixtures can be used with an acceptor molecule with a donor sugar, to produce *e.g.*, a polysaccharide, an oligosaccharide, a glycolipid, or a glycoprotein.

[0043] In another aspect, the invention provides a method of refolding an insoluble recombinant eukaryotic sialyltransferase, by (a) solubilizing the sialyltransferase; and then (b) contacting the soluble sialyltransferase with a refolding buffer including a redox couple. The refolded sialyltransferase is biologically active and catalyzes the transfer of sialic acid from a donor substrate to an acceptor substrate. In one embodiment, the refolded sialytransferase is dialyzed or diafiltered.

[0044] The refolding buffer can also include a detergent, or a chaotropic agent, or arginine. In some embodiments the pH of the refolding buffer is between 6.0 and 10.0. In one embodiment, the pH of the refolding buffer is between 6.5 and 8.0. In another embodiment, the pH of the refolding buffer is between 8.0 and 9.0. In another embodiment, the pH of the refolding buffer is between 7.5 and 8.5.

[0045] In one embodiment, the redox couple in the refolding buffer is reduced glutathione/oxidized glutathione (GSH/GSSG). In a further embodiment, the molar ratio of GSH/GSSG is between 100:1 and 1:10. In a preferred embodiment, the molar ratio of GSH/GSSG is 10:1. In a still further embodiment, the refolding buffer comprises about 0.02-

10 mM GSH, 0.005-10 mM GSSG, 0.005-10 mM lauryl maltoside, 50-250 mM NaCl, 2-10 mM KCl, 0.01-0.05% PEG 3350, and 150-550 mM L-arginine.

[0046] In another embodiment, the sialyltransferase includes an amino acid tag *e.g.*, maltose binding protein (MBP), a polyhistidine tag, a glutathione S transferase (GST), a starch binding protein (SBP), and a myc epitope. In a further embodiment, the sialyltransferase is purified using a tag binding molecule that binds to the amino acid tag. For example, the amino acid tag can be MBP and the tag binding molecule can be amylose, maltose, or a cyclodextrin.

[0047] In another embodiment, the refolded sialyltransferase catalyzes the transfer of sialic acid from CMP-sialic acid to a glycoprotein.

[0048] In a further embodiment, the refolded sialyltransferase catalyzes the transfer of 10KPEG or 20K PEG from CMP-SA-PEG(10 kDa) or CMP-SA-PEG(20 kDa) to a glycoprotein.

[0049] In another embodiment, the sialyltransferase is rat liver ST3GalIII.

[0050] In another aspect, the invention provides a method of adding a sialyl moiety to a glycoprotein, by contacting the glycoprotein with CMP-sialic acid with a refolded mammalian sialyltransferase that was refolded using the methods disclosed herein.

[0051] In another aspect, the invention provides a method of adding a PEG moiety to a glycoprotein, the method comprising by contacting the glycoprotein with CMP-SA-PEG(10 kDa) or CMP-SA-PEG(20 kDa) and a refolded mammalian sialyltransferase that was refolded using the methods disclosed herein.

[0052] In a further aspect the invention provides a method of refolding an insoluble recombinant eukaryotic N-acetylgalactosaminyltransferase 2 (GalNAcT2) by solubilizing the GalNAcT2 in a solubilization buffer; and then contacting the soluble GalNAcT2 with a refolding buffer that includes a redox couple to refold the GalNAcT2. After refolding, the refolded GalNAcT2 catalyzes the transfer of N-acetylgalactosamine from a donor substrate to an acceptor substrate. The method can optionally include steps of dialyzing or diafiltering the refolded GalNAcT2 or further purification of the refolded GalNAcT2.

[0053] In some embodiments the redox couple of the refolding buffer is reduced glutathione/oxidized glutathione (GSH/GSSG) or cysteine/cystamine. The refolding buffer

can also include the following: a detergent, a choatropic agent, or arginine. In some embodiments, the pH of the refolding buffer is between 6.0 and 10.0. In one preferred embodiment, the pH of the refolding buffer is about 8.0.

[0054] In preferred embodiments, the solubilization buffer pH is between 6.0 and 10.0. In a more preferred embodiment, the solubilization buffer pH is about 8.0.

[0055] The recombinantly expressed GalNAcT2 can include an amino acid tag. The amino acid tag can be, *e.g.*, a maltose binding protein (MBP), a polyhistidine tag, a glutathione S transferase (GST), a starch binding protein (SBP), or a myc epitope. A tag binding molecule can be used to purify the refolded GalNAcT2. When the amino acid tag is MBP and the tag binding molecule is generally one of the following: amylose, maltose, or a cyclodextrin.

[0056] In a preferred embodiment, the refolded GalNAcT2 catalyzes the transfer of N-acetylgalactosamine from a donor substrate to a peptide, protein, glycopeptide or glycoprotein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0057] Figure 1 provides the buffer conditions tested in refolding MBP-ST3GalIII from bacterial inclusion bodies. The activity of the refolded enzymes is also provided.

[0058] Figure 2 provides an elution profile of refolded and dialyzed MBP-ST3GalIII from an amylose column.

[0059] Figure 3 provides the ST3GalIII activities of the elution fractions from the amylose column.

[0060] Figure 4 provides the results of an assay of glycoPEGylation of transferrin using purified refolded MBP-ST3GalIII. Lanes are as follows: (1) MW markers [250, 148, 98, 64, 50 kD]; (2) Control asioalotransferrin with no enzyme, indicated by solid arrow; (3) transferrin-SA-PEG (20 kDa) production with Fraction #5, products indicated by arrowhead; (4) transferrin-SA-PEG (20 kDa) production with Fraction # 6, products indicated by arrowhead; (5) Purified, refolded MBP-ST3GalIII Fr # 6, indicated by dotted arrow; (6) MW markers; (7) same as 2; (8) transferrin-SA-PEG (10 kDa) production with Fr # 4, products indicated by brackets; and (9) transferrin-SA-PEG (10 kDa) production with Fr # 5, products indicated by brackets.

[0061] Figure 5 provides the results of an assay of GlycoPEGylation of EPO using the refolded SuperGlycoMix. Lanes are as follows: (1) MW markers, SeeBlue2 Invitrogen,(250, 148, 98, 64, 50, 36, 22, 16, 6 kD); (2) Positive control with EPO, + NSO expressed GalT1, BV GnT1, Aspergillus ST3GalIII and sugar nucleotides; (3) Negative control, Same as 2 without UDP-GlcNAc; (4) EPO, Purified and separately refolded MBP-GalT1(Δ129) C342T, Refolded MBP-GnT1(Δ103), and *Aspergillus niger* expressed ST3GalIII; (5) EPO, SuperGlycoMix (mixture of MBP-ST3GalIII, MBP-GalT1(Δ129) C342T, MBP-GnT1(Δ103)C123A and sugar nucleotides.

[0062] Figure 6 provides an alignment of a human GnT1 amino acid sequence (top line, NP_002397) and a rabbit GnT1 amino acid sequence (bottom line, P27115). The conserved unpaired cysteines are underlined and in bold text.

[0063] Figure 7 provides the amino acid sequence of a GnT1 Cys121Ser mutant and a nucleic acid sequence that encodes the mutant GnT1 protein. The amino acid sequence depicted begins with amino acid residue 104 of the full length human protein and is representative of mammalian GnT1 proteins with the following unpaired cysteine mutation: ...stvrrsldkllh...., where the bold residue is mutated from the wild-type cysteine.

[0064] Figure 8 provides the amino acid sequence of a GnT1 Cys121Asp mutant and a nucleic acid sequence that encodes the mutant GnT1 protein. The amino acid sequence depicted begins with amino acid residue 104 of the full length human protein and is representative of mammalian GnT1 proteins with the following unpaired cysteine mutation: ...stvrr**d**ldkllh..., where the bold residue is mutated from the wild-type cysteine.

[0065] Figure 9 provides the amino acid sequence of a GnT1 Cys121Thr mutant and a nucleic acid sequence that encodes the mutant GnT1 protein. The amino acid sequence depicted begins with amino acid residue 104 of the full length human protein and is representative of mammalian GnT1 proteins with the following unpaired cysteine mutation: ...stvrr**t**ldkllh..., where the bold residue is mutated from the wild-type cysteine.

[0066] Figure 10 provides the amino acid sequence of a GnT1 Cys121Ala mutant and a nucleic acid sequence that encodes the mutant GnT1 protein. The amino acid sequence depicted begins with amino acid residue 104 of the full length human protein and is representative of mammalian GnT1 proteins with the following unpaired cysteine mutation: ...stvrr**a**ldkllh..., where the bold residue is mutated from the wild-type cysteine.

[0067] Figure 11 provides the amino acid sequence of a GnT1 Arg120Ala, Cys121His mutant and a nucleic acid sequence that encodes the mutant GnT1 protein. The amino acid sequence depicted begins with amino acid residue 104 of the full length human protein and is representative of mammalian GnT1 proteins with the following double mutation: ...stvr**ah**ldkllh..., where the bold residue is mutated from the wild-type cysteine.

[0068] Figure 12 provides the amino acid sequence of rat liver ST3GalIII. The underlined and italicized sequence was deleted to make the Δ28 deletion.

[0069] Figures 13A and 13B provide full length nucleic acid and amino acid sequences of UDP-N-acetylgalactosaminyltransferase 2 (GalNAcT2). The accession number of the nucleic acid and protein is NM_004481.

[0070] Figures 14A and 14B provide nucleic acid and amino acid sequences of a Δ51GalNAcT2. The numbering is based on the full length amino acid and nucleic acid sequences shown in Figures 13A and B.

[0071] Figure 15 provides a demonstration of the protein concentration of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. The pH values tested are expressed as solubilization pH-refolding pH. Protein concentrations were measured immediately after refolding (light gray bars), after dialysis (dark gray bars), and after concentration (white bars).

[0072] Figure 16 provides a demonstration of the enzymatic activity of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. The pH values tested are expressed as solubilization pH-refolding pH. Activity was measured after dialysis (light gray bars) and after concentration (dark gray bars).

[0073] Figure 17 provides a demonstration of the specific activity of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. The pH values tested are expressed as solubilization pH-refolding pH. Specific activity was measured after dialysis (white bars) and after concentration (dark gray bars).

[0074] Figures 18A and 18B provide results of remodeling of recombinant granulocyte colony stimulating factor (GCSF) using refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. Figure 18A shows the results using a control purified MBP-GalNAcT2(D51), or a negative control that lacked a substrate, or

bacterially expressed MBP-GalNAcT2(D51) that was solubilized at pH 6.5 and refolded at pH 6.5. Figure 18B shows the experimental results.

[0075] Figure 19 provides a profile of refolded MBP-GalNAcT2(D51) proteins after elution from a Q Sepharose XL (QXL) column (Amersham Biosciences, Piscataway, NJ). The top of the figure shows a chromatogram illustrating the elution of MBP-GalNAcT2(D51) from the QXL column. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis. The bottom shows an image of two electrophoretic gels used to visualize the eluted fractions. The contents of each lane on the gel are described in the figure.

[0076] Figure 20 provides the GalNAcT2 activity of specific column fractions from the QXL column shown in Figure 19. The most active fractions were applied to a Hydroxyapatite Type I (80μm) (BioRad, Hercules, CA) column.

[0077] Figure 21 provides a profile of refolded MBP-GalNAcT2(D51) proteins after elution from the HA type I column. The top of the figure shows a chromatogram illustrating the elution of MBP-GalNAcT2(D51) from the HA type I column. Fraction numbers are indicated on the X-axis and the relative absorbance of each fraction is indicated on the Y-axis. The bottom shows an image of an electrophoretic gel used to visualize the eluted fractions. The contents of each lane on the gel are described in the figure.

[0078] Figure 22 provides the GalNAcT2 activity of HA type I eluted fractions.

[0079] Figure 23 provides a comparison of purification and activity of ST3Gal3 proteins fused to either an MBP tag or to an MBP-SBD tag.

[0080] Figure 24 provides the amino acid sequences of the MBP-ST3Gal1 fusion protein (A) and the MBP-SBD-ST3Gal1 fusion protein (B).

[0081] Figure 25 provides the sialyltransferase activity of the MBP-ST3Gal3 fusion protein) and the MBP-SBD-ST3Gal3 fusion protein. positive and negative controls are also shown.

[0082] Figure 26 provides the amino acid sequence of mouse and human ST6GalNAcI proteins fused to MBP. Part A shows the sequence of a mouse truncation fusion: MBP-mST6GalNAcI S127. Part B shows the sequence of a human truncation fusion: MBP-hST6GalNAcI K36.

[0083] Figure 27 provides SDS-PAGE gels of O-linked glycosyltransferase enzyme (A) concentrations after co-refolding and the (B) results of an enzyme assay after co-refolding. MBP-GalNAcT2 and MBP-ST3GalI were co-refolded together. Enzyme activity was tested after addition of Core I Gal T1 enzyme. The substrates were IFα-2b and 20K-Peg-CMP-NAN.

[0084] Figure 28 provides an SDS-PAGE gel showing expression of the native SiaA protein in *E. coli* before and after induction with IPTG.

[0085] Figure 29 provides an SDS-PAGE gel showing expression of an MBP-SiaA fusion protein in *E. coli* before and after induction with IPTG.

[0086] Figure 30 provides the amino acid sequence of the full length bovine GalT1 protein.

[0087] Figure 31 depicts GalT1 mutants schematically, as well as a control protein GalT1(40) (S96A+C342T).

[0088] Figure 32 provides the results of enzymatic assays of the refolded and purified MBP-GalT1 (D70) protein. The assay measured conversion of LNT2 (Lacto-N-Triose-2) into LNnT (Lacto-N-Neotetraose) using UDP-Gal (Uridine 5'-Diphosphogalactose) as a donor substrate.

[0089] Figure 33 provides an RNAse B remodeling assay of MBP-GalT1 (D70) and a control protein GalT1(40) (S96A+C342T), also referred to as Qasba's GalT1.

[0090] Figure 34 provides kinetics of glycosylation of RNAse B using the refolded and purified MBP-GalT1 (D70) protein or NSO GalT1, a soluble form of the bovine GalT1 protein that was expressed in a mammalian cell system.

[0091] Figure 35 provides a schematic of the MBP-GnT1 fusion proteins, and depicts the truncations, *e.g.*, Δ103 or Δ35, and the Cys121Ser mutation (top). The bottom of the figure provides the full length human GnT1 protein.

[0092] Figure 36 provides an SDS-PAGE gel showing in the right panel the refolded MBP-GnT1 fusion proteins: MBP-GnT1(D35) C121A, MBP-GnT1(D103) R120A + C121H, and MBP-GnT1(D103) C121A. The left panel shows GnT1 activities of two different batches (A1 and A2) of refolded MBP-GnT1(D35) C121A at different time points.

[0093] Figure 37 provides a full length sequence of porcine ST3Gal1.

[0094] Figure 38 provides full length amino acid sequences for A) human ST6GalNAcI and for B) chicken ST6GalNAcI, and C) a sequence of the mouse ST6GalNAcI protein beginning at residue 32 of the native mouse protein.

[0095] Figure 39 provides a schematic of a number of preferred human ST6GalNAcI truncation mutants.

[0096] Figure 40 shows a schematic of MBP fusion proteins including the human ST6GalNAcI truncation mutants.

[0097] Figure 41 provides the full length sequence of human Core 1 GalT1 protein.

[0098] Figure 42 provides the sequences of two Drosophila Core 1 GalT1 proteins.

[0099] Figure 43 provides the sequences of exemplary bacterial MBP proteins that can be fused to glycosyltransferases to enhance refolding. A. *Yersinia* MBP; B. *E. coli* MBP; C. *Pyrococcus furiosus* MBP; D. *Thermococcus litoralis* MBP; E. *Thermatoga maritime* MBP; and F. *Vibrio cholerae* MBP.

[0100] Figure 44 provides an alignment of human GalNAcT1 and GalNAcT2 proteins. Because the alignment programs account for sequence insertions or deletions, the numbering of cysteine residues is not the same as mentioned text and published sequences. In the case of hGalNAc-T2 cysteine 227 (published) corresponds to position 235 in the alignment and cysteine 229 (published) is 237 in the alignment. The hGalNAc-T1 cysteines are 212 (published), which corresponds to cysteine 235 (alignment) and 214 (published), which corresponds to cysteine 237 (alignment). The relevant cysteine residues are indicated by larger font size.

[0101] Figure 45 shows the position of paired and unpaired cysteine residues in the human ST6GalNAcI protein. Single and double cysteine substitution are also shown, *e.g.*, C280S, C362S, C362T, (C280S + C362S), and (C280S + C362T).

DEFINITIONS

[0102] The recombinant glycosyltransferase proteins of the invention are useful for transferring a saccharide from a donor substrate to an acceptor substrate. The addition generally takes place at the non-reducing end of an oligosaccharide or carbohydrate moiety on a biomolecule. Biomolecules as defined here include but are not limited to biologically

significant molecules such as carbohydrates, proteins (*e.g.*, glycoproteins), and lipids (*e.g.*, glycolipids, phospholipids, sphingolipids and gangliosides).

The following abbreviations are used herein:

Ara = arabinosyl;

Fru = fructosyl;

Fuc = fucosyl;

Gal = galactosyl;

GalNAc = N-acetylgalactosylamino;

Glc = glucosyl;

GlcNAc = N-acetylglucosylamino;

Man = mannosyl; and

NeuAc = sialyl (N-acetylneuraminyl)

FT or FucT = fucosyltransferase*

ST = sialyltransferase*

GalT = galactosyltransferase*

[0103] Arabic or Roman numerals are used interchangeably herein according to the naming convention used in the art to indicate the identity of a specific glycosyltransferase (*e.g.*, FTVII and FT7 refer to the same fucosyltransferase).

[0104] Oligosaccharides are considered to have a reducing end and a non-reducing end, whether or not the saccharide at the reducing end is in fact a reducing sugar. In accordance with accepted nomenclature, oligosaccharides are depicted herein with the non-reducing end on the left and the reducing end on the right.

[0105] All oligosaccharides described herein are described with the name or abbreviation for the non-reducing saccharide (*e.g.*, Gal), followed by the configuration of the glycosidic bond (α or β), the ring bond, the ring position of the reducing saccharide involved in the bond, and then the name or abbreviation of the reducing saccharide (*e.g.*, GlcNAc). The linkage between two sugars may be expressed, for example, as 2,3, 2→3, or (2,3). Each saccharide is a pyranose or furanose.

[0106] The term "sialic acid" refers to any member of a family of nine-carbon carboxylated sugars. The most common member of the sialic acid family is N-acetyl-neuraminic acid (2-keto-5-acetamido-3,5-dideoxy-D-glycero-D-galactononulopyranos-1-onic acid (often abbreviated as Neu5Ac, NeuAc, or NANA). A second member of the family is N-glycolyl-

neuraminic acid (Neu5Gc or NeuGc), in which the N-acetyl group of NeuAc is hydroxylated. A third sialic acid family member is 2-keto-3-deoxy-nonulosonic acid (KDN) (Nadano *et al.* (1986) *J. Biol. Chem.* **261**: 11550-11557; Kanamori *et al.*, *J. Biol. Chem.* **265**: 21811-21819 (1990)). Also included are 9-substituted sialic acids such as a 9-O-C_1-C_6 acyl-Neu5Ac like 9-O-lactyl-Neu5Ac or 9-O-acetyl-Neu5Ac, 9-deoxy-9-fluoro-Neu5Ac and 9-azido-9-deoxy-Neu5Ac. For review of the sialic acid family, *see, e.g.*, Varki, *Glycobiology* **2**: 25-40 (1992); *Sialic Acids: Chemistry, Metabolism and Function*, R. Schauer, Ed. (Springer-Verlag, New York (1992)). The synthesis and use of sialic acid compounds in a sialylation procedure is disclosed in international application WO 92/16640, published October 1, 1992.

[0107] An "acceptor substrate" for a glycosyltransferase is an oligosaccharide moiety that can act as an acceptor for a particular glycosyltransferase. When the acceptor substrate is contacted with the corresponding glycosyltransferase and sugar donor substrate, and other necessary reaction mixture components, and the reaction mixture is incubated for a sufficient period of time, the glycosyltransferase transfers sugar residues from the sugar donor substrate to the acceptor substrate. The acceptor substrate will often vary for different types of a particular glycosyltransferase. For example, the acceptor substrate for a mammalian galactoside 2-L-fucosyltransferase (α1,2-fucosyltransferase) will include a Galβ1,4-GlcNAc-R at a non-reducing terminus of an oligosaccharide; this fucosyltransferase attaches a fucose residue to the Gal via an α1,2 linkage. Terminal Galβ1,4-GlcNAc-R and Galβ1,3-GlcNAc-R and sialylated analogs thereof are acceptor substrates for α1,3 and α1,4-fucosyltransferases, respectively. These enzymes, however, attach the fucose residue to the GlcNAc residue of the acceptor substrate. Accordingly, the term "acceptor substrate" is taken in context with the particular glycosyltransferase of interest for a particular application. Acceptor substrates for additional glycosyltransferases, are described herein. Acceptor substrates also include *e.g.*, peptides, proteins, glycopeptides, and glycoproteins.

[0108] A "donor substrate" for glycosyltransferases is an activated nucleotide sugar. Such activated sugars generally consist of uridine, guanosine, and cytidine monophosphate derivatives of the sugars (UMP, GMP and CMP, respectively) or diphosphate derivatives of the sugars (UDP, GDP and CDP, respectively) in which the nucleoside monophosphate or diphosphate serves as a leaving group. For example, a donor substrate for fucosyltransferases is GDP-fucose. Donor substrates for sialyltransferases, for example, are activated sugar nucleotides comprising the desired sialic acid. For instance, in the case of NeuAc, the activated sugar is CMP-NeuAc. Other donor substrates include *e.g.*, GDP mannose, UDP-

galactose, UDP-*N*-acetylgalactosamine, CMP-NeuAc-PEG (also referred to as CMP-sialic acid-PEG), UDP-*N*-acetylglucosamine, UDP-glucose, UDP-glucorionic acid, and UDP-xylose. Sugars include, *e.g.*, NeuAc, mannose, galactose, *N*-acetylgalactosamine, *N*-acetylglucosamine, glucose, glucorionic acid, and xylose. Bacterial, plant, and fungal systems can sometimes use other activated nucleotide sugars.

[0109] A " method of remodeling a protein, a peptide, a glycoprotein, or a glycopeptide" as used herein, refers to addition of a sugar residue to a protein, a peptide, a glycoprotein, or a glycopeptide using a glycosyltransferase. In a preferred embodiment, the sugar residue is covalently attached to a PEG molecule.

[0110] A "eukaryotic glycosyltransferase" as used herein refers to an enzyme that is derived from a eukaryotic organism and that catalyzes transfer of a sugar reside from a donor substrate, *i.e.*, an activated nucleotide sugar to an acceptor substrate, *e.g.*, an oligosaccharide, a glycolipid, a peptide, a protein, a glycopeptide, or a glycoprotein. In preferred embodiments, a eukaryotic glycosyltransferase transfers a sugar from a donor substrate to a peptide, a protein, a glycopeptide, or a glycoprotein. In another preferred embodiment, a eukaryotic glycosyltransferase is a type II transmembrane glycosyltransferase. A eukaryotic glycosyltransferase can be derived from an eukaryotic organism, *e.g.*, a multicellular eukaryotic organism, a plant, an invertebrate animal, such as *Drosophila* or *C. elegans*, a vertebrate animal, an amphibian or reptile, a mammal, a rodent, a primate, a human, a rabbit, a rat, a mouse, a cow, or a pig and so on.

[0111] A "eukaryotic *N*-acetylglucosaminyltransferase I (GnTI or GNTI)" as used herein, refers to a β-1,2-*N*- acetylglucosaminyltransferase I isolated from a eukaryotic organism. The enzyme catalyzes the transfer of *N*-acetylglucosamine (GlcNAc) from a UDP-GlcNAc donor to an acceptor molecule comprising a mannose sugar. Like other eukaryotic glycosyltransferases, GnTI has a transmembrane domain, a stem region, and a catalytic domain. Eukaryotic GnT1 proteins include, *e.g.*, human, accession number NP_002397; Chinese hamster, accession number AAK61868; rabbit, accession number AAA31493; rat, accession number NP_110488; golden hamster, accession number AAD04130; mouse, accession number P27808; zebrafish, accession number AAH58297; *Xenopus*, accession number CAC51119; *Drosophila*, accession number NP_525117; *Anopheles*, accession number XP_315359; *C. elegans*, accession number NP_497719; *Physcomitrella patens*, accession number CAD22107; *Solanum tuberosum*, accession number CAC80697; *Nicotiana*

tabacum, accession number CAC80702; *Oryza sativa*, accession number CAD30022; *Nicotiana benthamiana*, accession number CAC82507; and *Arabidopsis thaliana*, accession number NP_195537, each of which are herein incorporated by reference.

[0112] A "eukaryotic *N*-acetylgalactosaminyltransferase (GalNAcT)" as used herein, refers to an *N*-acetylgalactosaminyltransferase isolated from a eukaryotic organism. The enzyme catalyzes the transfer of *N*-acetylgalactosamine (GalNAc) from a UDP-GalN. donor to an acceptor molecule. Like other eukaryotic glycosyltransferases, GalNAcT enzymes have a transmembrane domain, a stem region, and a catalytic domain. A number of GalNAcT enzymes have been isolated and characterized, *e.g.*, GalNAcT1, accession number X85018; GalNAcT2, accession number X85019 (both described in White *et al.*, *J. Biol. Chem.* 270:24156-24165 (1995)); and GalNAcT3, accession number X92689 (described in Bennett *et al.*, *J. Biol. Chem.* 271:17006-17012 (1996), each of which are herein incorporated by reference).

[0113] A "eukaryotic β-1,4-galactosyltransferase (GalT1) as used herein, refers to a β-1,4-galactosyltransferase isolated from a eukaryotic organism. The enzyme catalyzes the transfer of galactose from a UDP-Gal donor to an acceptor molecule. Like other eukaryotic glycosyltransferases, GalT1 enzymes have a transmembrane domain, a stem region, and a catalytic domain. A number of GalT1 enzymes have been isolated and characterized, *e.g.*, the full length bovine sequence, D'Agostaro *et al.*, *Eur. J. Biochem.* 183:211-217 (1989) and accession number CAA32695, each of which are herein incorporated by reference.

[0114] A "eukaryotic α(2,3)sialyltransferase (ST3Gal3)" as used herein, refers to an α(2,3)sialyltransferase isolated from a eukaryotic organism. This enzyme catalyzes the transfer of sialic acid to the Gal of a Galβ1,3GlcNAc, Galβ1,3GalNAc or Galβ1,4GlcNAc glycoside (*see, e.g.*, Wen *et al.* (1992) *J. Biol. Chem.* **267**: 21011; Van den Eijnden *et al.* (1991) *J. Biol. Chem.* **256:** 3159). The sialic acid is linked to a Gal with the formation of an α-linkage between the two saccharides. Bonding (linkage) between the saccharides is between the 2-position of NeuAc and the 3-position of Gal. Like other eukaryotic glycosyltransferases, ST3GalIII enzymes have a transmembrane domain, a stem region, and a catalytic domain. This particular enzyme can be isolated from rat liver (Weinstein *et al.* (1982) *J. Biol. Chem.* **257**: 13845); the human cDNA (Sasaki *et al.* (1993) *J. Biol. Chem.* **268**: 22782-22787; Kitagawa & Paulson (1994) *J. Biol. Chem.* **269**: 1394-1401) and genomic (Kitagawa *et al.* (1996) *J. Biol. Chem.* **271**: 931-938) DNA sequences are known, facilitating

production of this enzyme by recombinant expression. Rat ST3GalIII has been cloned and the sequence is known. *See, e.g.*, Wen *et al.*, *J. Biol. Chem.* 267:21011-21019 (1992) and Accession number M97754, each of which are herein incorporated by reference.

[0115] A "eukaryotic α-N-acetylgalactosaminide α-2,6-sialyltransferase I (ST6GalNAcT1) as used herein, refers to an α(2,6)sialyltransferase isolated from a eukaryotic organism. The enzyme catalyzes the transfer of sialic acid from a CMP-sialic acid donor to an acceptor molecule. The transfer is an α2,6-linkage to N-acetylgalactosamine-O-Thr/Ser. Like other eukaryotic glycosyltransferases, ST6GalNAcT1 enzymes have a transmembrane domain, a stem region, and a catalytic domain. A number of ST6GalNAcT1 enzymes have been isolated and characterized, *e.g.*, the full length mouse sequence, Kurosawa *et al.*, *J. Biochem.* 127:845-854 (2000) and accession number JC7248, each of which are herein incorporated by reference. Other exemplary ST6GalNAcT1 amino acid sequences are found in Figure 38.

[0116] A "eukaryotic gal β1,3GalNAc α2,3-sialyltransferase (ST3GalI)" as used herein, refers to a gal β1,3GalNAc α2,3-sialyltransferase isolated from a eukaryotic organism. The enzyme catalyzes the transfer of sialic acid from a CMP-sialic acid donor to an acceptor molecule. The transfer is an α2,3-linkage to N-acetylgalactosamine-O-Thr/Ser. Like other eukaryotic glycosyltransferases, ST3GalI enzymes have a transmembrane domain, a stem region, and a catalytic domain. A number of ST3GalI enzymes have been isolated and characterized, *e.g.*, the full length porcine sequence, Gillespie *et al.*, *J. Biol.* Chem. 267:21004-21010 (1992) and accession number A45073, each of which are herein incorporated by reference.

[0117] A "eukaryotic core I galactosyltransferase (Core 1 GalT1)" as used herein refers to a protein with Core 1 β1,3-Galactosyltransferase activity. Like other eukaryotic glycosyltransferases, Core 1 GalT1 enzymes have a transmembrane domain, a stem region, and a catalytic domain. A number of Core 1 GalT1 enzymes have been isolated and characterized, *e.g.*, the Drosophila and human sequences of Figures 41 and 42. The human protein is characterized in Ju *et al.*, *J. Biol. Chem.* 277 (1), 178-186 (2002), which is herein incorporated by reference for all purposes.

[0118] An "unpaired cysteine residue" as used herein, refers to a cysteine residue, which in a correctly folded protein (*i.e.*, a protein with biological activity), does not form a disulfide bind with another cysteine residue.

[0119] An "insoluble glycosyltransferase" refers to a glycosyltransferase that is expressed in bacterial inclusion bodies. Insoluble glycosyltransferases are typically solubilized or denatured using *e.g.*, detergents or chaotropic agents or some combination. "Refolding" refers to a process of restoring the structure of a biologically active glycosyltransferase to a glycosyltransferase that has been solubilized or denatured. Thus, a refolding buffer, refers to a buffer that enhances or accelerates refolding of a glycosyltransferase.

[0120] A "redox couple" refers to mixtures of reduced and oxidized thiol reagents and include reduced and oxidized glutathione (GSH/GSSG), cysteine/cystine, cysteamine/cystamine, DTT/GSSG, and DTE/GSSG. (*See, e.g.*, Clark, *Cur. Op. Biotech.* 12:202-207 (2001)).

[0121] The term "contacting" is used herein interchangeably with the following: combined with, added to, mixed with, passed over, incubated with, flowed over, etc.

[0122] The term "PEG" refers to poly(ethylene glycol). PEG is an exemplary polymer that has been conjugated to peptides. The use of PEG to derivatize peptide therapeutics has been demonstrated to reduce the immunogenicity of the peptides and prolong the clearance time from the circulation. For example, U.S. Pat. No. 4,179,337 (Davis *et al.*) concerns non-immunogenic peptides, such as enzymes and peptide hormones coupled to polyethylene glycol (PEG) or polypropylene glycol. Between 10 and 100 moles of polymer are used per mole peptide and at least 15% of the physiological activity is maintained.

[0123] The term "specific activity" as used herein refers to the catalytic activity of an enzyme, *e.g.*, a recombinant glycosyltransferase fusion protein of the present invention, and may be expressed in activity units. As used herein, one activity unit catalyzes the formation of 1 μmol of product per minute at a given temperature (*e.g.*, at 37°C) and pH value (*e.g.*, at pH 7.5). Thus, 10 units of an enzyme is a catalytic amount of that enzyme where 10 μmol of substrate are converted to 10 μmol of product in one minute at a temperature of, *e.g.*, 37 °C and a pH value of, *e.g.*, 7.5.

[0124] "N-linked" oligosaccharides are those oligosaccharides that are linked to a peptide backbone through asparagine, by way of an asparagine-N-acetylglucosamine linkage. N-linked oligosaccharides are also called "N-glycans." All N-linked oligosaccharides have a common pentasaccharide core of $Man_3GlcNAc_2$. They differ in the presence of, and in the number of branches (also called antennae) of peripheral sugars such as N-acetylglucosamine,

galactose, N-acetylgalactosamine, fucose and sialic acid. Optionally, this structure may also contain a core fucose molecule and/or a xylose molecule.

[0125] "O-linked" oligosaccharides are those oligosaccharides that are linked to a peptide backbone through threonine, serine, hydroxyproline, tyrosine, or other hydroxy-containing amino acids.

[0126] A "substantially uniform glycoform" or a "substantially uniform glycosylation pattern," when referring to a glycoprotein species, refers to the percentage of acceptor substrates that are glycosylated by the glycosyltransferase of interest (*e.g.*, fucosyltransferase). It will be understood by one of skill in the art, that the starting material may contain glycosylated acceptor substrates. Thus, the calculated amount of glycosylation will include acceptor substrates that are glycosylated by the methods of the invention, as well as those acceptor substrates already glycosylated in the starting material.

[0127] The term "biological activity" refers to an enzymatic activity of a protein. For example, biological activity of a sialyltransferase refers to the activity of transferring a sialic acid moiety from a donor molecule to an acceptor molecule. Biological activity of a GalNAcT2 refers to the activity of transferring an *N*-acetylgalactosamine moiety from a donor molecule to an acceptor molecule. For GalNAcT2 proteins, an acceptor molecule can be a protein, a peptide, a glycoprotein, or a glycopeptide. Biological activity of a GnT1 protein refers to the activity of transferring a *N*-acetylglucosamine moiety from a donor molecule to an acceptor molecule. Biological activity of a galactosyltransferase refers to the activity of transferring a galactose moiety from a donor molecule to an acceptor molecule.

[0128] "Commercial scale" refers to gram scale production of a product saccharide in a single reaction. In preferred embodiments, commercial scale refers to production of greater than about 50, 75, 80, 90 or 100, 125, 150, 175, or 200 grams.

[0129] The term "substantially" in the above definitions of "substantially uniform" generally means at least about 60%, at least about 70%, at least about 80%, or more preferably at least about 90%, and still more preferably at least about 95% of the acceptor substrates for a particular glycosyltransferase are glycosylated.

[0130] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the

genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

[0131] "Protein", "polypeptide", or "peptide" refer to a polymer in which the monomers are amino acids and are joined together through amide bonds, alternatively referred to as a polypeptide. When the amino acids are α-amino acids, either the L-optical isomer or the D-optical isomer can be used. Additionally, unnatural amino acids, for example, β-alanine, phenylglycine and homoarginine are also included. Amino acids that are not gene-encoded may also be used in the present invention. Furthermore, amino acids that have been modified to include reactive groups may also be used in the invention. All of the amino acids used in the present invention may be either the D - or L -isomer. The L -isomers are generally preferred. In addition, other peptidomimetics are also useful in the present invention. For a general review, *see,* Spatola, A. F., in CHEMISTRY AND BIOCHEMISTRY OF AMINO ACIDS, PEPTIDES AND PROTEINS, B. Weinstein, eds., Marcel Dekker, New York, p. 267 (1983).

[0132] The term "recombinant" when used with reference to a cell indicates that the cell replicates a heterologous nucleic acid, or expresses a peptide or protein encoded by a heterologous nucleic acid. Recombinant cells can contain genes that are not found within the native (non-recombinant) form of the cell. Recombinant cells can also contain genes found in the native form of the cell wherein the genes are modified and re-introduced into the cell by artificial means. The term also encompasses cells that contain a nucleic acid endogenous to the cell that has been modified without removing the nucleic acid from the cell; such modifications include those obtained by gene replacement, site-specific mutation, and related techniques. A "recombinant protein" is one which has been produced by a recombinant cell. In preferred embodiments, a recombinant eukaryotic glycosyltransferase is produced by a recombinant bacterial cell.

[0133] A "fusion protein" refers to a protein comprising amino acid sequences that are in addition to, in place of, less than, and/or different from the amino acid sequences encoding the original or native full-length protein or subsequences thereof.

[0134] Components of fusion proteins include "accessory enzymes" and/or "purification tags." An "accessory enzyme" as referred to herein, is an enzyme that is involved in catalyzing a reaction that, for example, forms a substrate for a glycosyltransferase. An accessory enzyme can, for example, catalyze the formation of a nucleotide sugar that is used as a donor moiety by a glycosyltransferase. An accessory enzyme can also be one that is used in the generation of a nucleotide triphosphate required for formation of a nucleotide sugar, or in the generation of the sugar which is incorporated into the nucleotide sugar. The recombinant fusion protein of the invention can be constructed and expressed as a fusion protein with a molecular "purification tag" at one end, which facilitates purification of the protein. Such tags can also be used for immobilization of a protein of interest during the glycosylation reaction. Suitable tags include "epitope tags," which are a protein sequence that is specifically recognized by an antibody. Epitope tags are generally incorporated into fusion proteins to enable the use of a readily available antibody to unambiguously detect or isolate the fusion protein. A "FLAG tag" is a commonly used epitope tag, specifically recognized by a monoclonal anti-FLAG antibody, consisting of the sequence AspTyrLysAspAspAsp AspLys or a substantially identical variant thereof. Other suitable tags are known to those of skill in the art, and include, for example, an affinity tag such as a hexahistidine peptide, which will bind to metal ions such as nickel or cobalt ions. Proteins comprising purification tags can be purified using a binding partner that binds the purification tag, *e.g.*, antibodies to the purification tag, nickel or cobalt ions or resins, and amylose, maltose, or a cyclodextrin. Purification tags also include starch binding domains, *E. coli* thioredoxin domains (vectors and antibodies commercially available from *e.g.*, Santa Cruz Biotechnology, Inc. and Alpha Diagnostic International, Inc.), and the carboxy-terminal half of the SUMO protein (vectors and antibodies commercially available from *e.g.*, Life Sensors Inc.). Maltose binding domains are preferably used for their ability to enhance refolding of insoluble eukaryotic glycosyltransferases, but can also be used to assist in purification of a fusion protein. Purification of maltose binding domain proteins is known to those of skill in the art. Starch binding domains are described in WO 99/15636, herein incorporated by reference. Affinity purification of a fusion protein comprising a starch binding domain using

a betacyclodextrin (BCD)-derivatized resin is described in USSN 60/468,374, filed May 5, 2003, herein incorporated by reference in its entirety.

[0135] The term "functional domain" with reference to glycosyltransferases, refers to a domain of the glycosyltransferase that confers or modulates an activity of the enzyme, *e.g.*, acceptor substrate specificity, catalytic activity, binding affinity, localization within the Golgi apparatus, anchoring to a cell membrane, or other biological or biochemical activity. Examples of functional domains of glycosyltransferases include, but are not limited to, the catalytic domain, stem region, and signal-anchor domain.

[0136] The terms "expression level" or "level of expression" with reference to a protein refers to the amount of a protein produced by a cell. The amount of protein produced by a cell can be measured by the assays and activity units described herein or known to one skilled in the art. One skilled in the art would know how to measure and describe the amount of protein produced by a cell using a variety of assays and units, respectively. Thus, the quantitation and quantitative description of the level of expression of a protein, *e.g.*, a glycosyltransferase, is not limited to the assays used to measure the activity or the units used to describe the activity, respectively. The amount of protein produced by a cell can be determined by standard known assays, for example, the protein assay by Bradford (1976), the bicinchoninic acid protein assay kit from Pierce (Rockford, Illinois), or as described in U.S. Patent No. 5,641,668.

[0137] The term "enzymatic activity" refers to an activity of an enzyme and may be measured by the assays and units described herein or known to one skilled in the art. Examples of an activity of a glycosyltransferase include, but are not limited to, those associated with the functional domains of the enzyme, *e.g.*, acceptor substrate specificity, catalytic activity, binding affinity, localization within the Golgi apparatus, anchoring to a cell membrane, or other biological or biochemical activity.

[0138] A "stem region" with reference to glycosyltransferases refers to a protein domain, or a subsequence thereof, which in the native glycosyltransferases is located adjacent to the trans-membrane domain, and has been reported to function as a retention signal to maintain the glycosyltransferase in the Golgi apparatus and as a site of proteolytic cleavage. Stem regions generally start with the first hydrophilic amino acid following the hydrophobic transmembrane domain and end at the catalytic domain, or in some cases the first cysteine residue following the transmembrane domain. Exemplary stem regions include, but is not

limited to, the stem region of fucosyltransferase VI, amino acid residues 40-54; the stem region of mammalian GnT1, amino acid residues from about 36 to about 103 (see, *e.g.*, the human enzyme); the stem region of mammalian GalT1, amino acid residues from about 71 to about 129 (see *e.g.*, the bovine enzyme); the stem region of mammalian ST3GalIII, amino acid residues from about 29 to about 84 (see, *e.g.*, the rat enzyme); the stem region of invertebrate Core 1GalT1, amino acid residues from about 36 to about 102 (see *e.g.*, the *Drosophila* enzyme); the stem region of mammalian Core 1 GalT1, amino acid residues from about 32 to about 90 (see *e.g.*, the human enzyme); the stem region of mammalian ST3Gal1, amino acid residues from about 28 to about 61 (see *e.g.*, the porcine enzyme) or for the human enzyme amino acid residues from about 18 to about 58; the stem region of mammalian ST6GalNAcI, amino acid residues from about 30 to about 207 (see *e.g.*, the murine enzyme), amino acids 35-278 for the h uman enzyme or amino acids 37-253 for the chicken enzyme; the stem region of mammalian GalNAcT2, amino acid residues from about 71 to about 129 (see *e.g.*, the rat enzyme).

[0139] A "catalytic domain" refers to a protein domain, or a subsequence thereof, that catalyzes an enzymatic reaction performed by the enzyme. For example, a catalytic domain of a sialyltransferase will include a subsequence of the sialyltransferase sufficient to transfer a sialic acid residue from a donor to an acceptor saccharide. A catalytic domain can include an entire enzyme, a subsequence thereof, or can include additional amino acid sequences that are not attached to the enzyme, or a subsequence thereof, as found in nature. An exemplary catalytic region is, but is not limited to, the catalytic domain of fucosyltransferase VII, amino acid residues 39-342; the catalytic domain of mammalian GnT1, amino acid residues from about 104 to about 445 (see, *e.g.*, the human enzyme); the catalytic domain of mammalian GalT1, amino acid residues from about 130 to about 402 (see *e.g.*, the bovine enzyme); and the catalytic domain of mammalian ST3GalIII, amino acid residues from about 85 to about 374 (see, *e.g.*, the rat enzyme). Catalytic domains and truncation mutants of GalNAcT2 proteins are described in USSN 60/576,530 filed June 3, 2004; and US provisional patent application Attorney Docket Number 040853-01-5149-P1, filed August 3, 2004; both of which are herein incorporated by reference for all purposes. Catalytic domains can also be identified by alignment with known glycosyltransferases.

[0140] A "subsequence" refers to a sequence of nucleic acids or amino acids that comprise a part of a longer sequence of nucleic acids or amino acids (*e.g.*, protein) respectively.

[0141] A "glycosyltransferase truncation" or a "truncated glycosyltransferase" or grammatical variants, refer to a glycosyltransferase that has fewer amino acid residues than a naturally occurring glycosyltransferase, but that retains enzymatic activity. Truncated glycosyltransferases include, *e.g.*, truncated GnT1 enzymes, truncated GalT1 enzymes, truncated ST3GalIII enzymes, truncated GalNAcT2 enzymes, truncated Core 1 GalT1 enzymes, amino acid residues from about 32 to about 90 (see *e.g.*, the human enzyme); truncated ST3Gal1 enzymes, truncated ST6GalNAcI enzymes, and truncated GalNAcT2 enzymes. Any number of amino acid residues can be deleted so long as the enzyme retains activity. In some embodiments, domains or portions of domains can be deleted, *e.g.*, a signal-anchor domain can be deleted leaving a truncation comprising a stem region and a catalytic domain; a signal-anchor domain and a portion of a stem region can be deleted leaving a truncation comprising the remaining stem region and a catalytic domain; or a signal-anchor domain and a stem region can be deleted leaving a truncation comprising a catalytic domain.

[0142] The term "nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single-or double-stranded form, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof.

[0143] A "recombinant expression cassette" or simply an "expression cassette" is a nucleic acid construct, generated recombinantly or synthetically, with nucleic acid elements that are capable of affecting expression of a structural gene in hosts compatible with such sequences. Expression cassettes include at least promoters and optionally, transcription termination signals. Typically, the recombinant expression cassette includes a nucleic acid to be transcribed (*e.g.*, a nucleic acid encoding a desired polypeptide), and a promoter. Additional factors necessary or helpful in effecting expression may also be used as described herein. For example, an expression cassette can also include nucleotide sequences that encode a signal sequence that directs secretion of an expressed protein from the host cell. Transcription termination signals, enhancers, and other nucleic acid sequences that influence gene expression, can also be included in an expression cassette. In preferred embodiments, a recombinant expression cassette encoding an amino acid sequence comprising a eukaryotic glycosyltransferase is expressed in a bacterial host cell.

[0144] A "heterologous sequence" or a "heterologous nucleic acid", as used herein, is one that originates from a source foreign to the particular host cell, or, if from the same source, is modified from its original form. Thus, a heterologous glycoprotein gene in a eukaryotic host cell includes a glycoprotein-encoding gene that is endogenous to the particular host cell that has been modified. Modification of the heterologous sequence may occur, *e.g.*, by treating the DNA with a restriction enzyme to generate a DNA fragment that is capable of being operably linked to the promoter. Techniques such as site-directed mutagenesis are also useful for modifying a heterologous sequence.

[0145] The term "isolated" refers to material that is substantially or essentially free from components which interfere with the activity of an enzyme. For a saccharide, protein, or nucleic acid of the invention, the term "isolated" refers to material that is substantially or essentially free from components which normally accompany the material as found in its native state. Typically, an isolated saccharide, protein, or nucleic acid of the invention is at least about 80% pure, usually at least about 90%, and preferably at least about 95% pure as measured by band intensity on a silver stained gel or other method for determining purity. Purity or homogeneity can be indicated by a number of means well known in the art. For example, a protein or nucleic acid in a sample can be resolved by polyacrylamide gel electrophoresis, and then the protein or nucleic acid can be visualized by staining. For certain purposes high resolution of the protein or nucleic acid may be desirable and HPLC or a similar means for purification, for example, may be utilized.

[0146] The term "operably linked" refers to functional linkage between a nucleic acid expression control sequence (such as a promoter, signal sequence, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence affects transcription and/or translation of the nucleic acid corresponding to the second sequence.

[0147] The terms "identical" or percent "identity," in the context of two or more nucleic acids or protein sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection.

[0148] The phrase "substantially identical," in the context of two nucleic acids or proteins, refers to two or more sequences or subsequences that have at least greater than about 60%

nucleic acid or amino acid sequence identity, 65%, 70%, 75%, 80%, 85%, 90%, preferably 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. Preferably, the substantial identity exists over a region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably the sequences are substantially identical over at least about 150 residues. In a most preferred embodiment, the sequences are substantially identical over the entire length of the coding regions.

[0149] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[0150] Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (*see generally*, *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1995 Supplement) (Ausubel)).

[0151] Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (1990) *J. Mol. Biol.* 215: 403-410 and Altschuel *et al.* (1977) *Nucleic Acids Res.* 25: 3389-3402, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the

same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al, supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

[0152] In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

[0153] A further indication that two nucleic acid sequences or proteins are substantially identical is that the protein encoded by the first nucleic acid is immunologically cross reactive with the protein encoded by the second nucleic acid, as described below. Thus, a protein is typically substantially identical to a second protein, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions, as described below.

[0154] The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) DNA or RNA.

[0155] The term "stringent conditions" refers to conditions under which a probe will hybridize to its target subsequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 15°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. (As the target sequences are generally present in excess, at Tm, 50% of the probes are occupied at equilibrium). Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60°C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is typically at least two times background, preferably 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5x SSC, and 1% SDS, incubating at 42° C, or, 5x SSC, 1% SDS, incubating at 65° C, with wash in 0.2x SSC, and 0.1% SDS at 65° C. For PCR, a temperature of about 36° C is typical for low stringency amplification, although annealing temperatures may vary between about 32-48° C depending on primer length. For high stringency PCR amplification, a temperature of about 62° C is typical, although high stringency annealing temperatures can range from about 50° C to about 65° C, depending on the primer length and specificity. Typical cycle conditions for both high and low stringency amplifications include a denaturation phase of 90-95° C for 30-120 sec, an annealing phase lasting 30-120 sec, and an extension phase of about 72° C for 1-2 min. Protocols and guidelines for low and high stringency amplification reactions are available, e.g., in Innis, et al. (1990) *PCR Protocols: A Guide to Methods and Applications* Academic Press, N.Y.

[0156] The phrases "specifically binds to a protein" or "specifically immunoreactive with", when referring to an antibody refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind preferentially to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to a protein under such conditions requires an antibody that is selected for its specificity for a particular protein. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York, for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity.

[0157] "Conservatively modified variations" of a particular polynucleotide sequence refers to those polynucleotides that encode identical or essentially identical amino acid sequences, or where the polynucleotide does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons CGU, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded protein. Such nucleic acid variations are "silent variations," which are one species of "conservatively modified variations." Every polynucleotide sequence described herein which encodes a protein also describes every possible silent variation, except where otherwise noted. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and UGG which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each "silent variation" of a nucleic acid which encodes a protein is implicit in each described sequence.

[0158] Furthermore, one of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations" where the alterations result in the substitution of an

amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.

[0159] One of skill will appreciate that many conservative variations of proteins, *e.g.*, glycosyltransferases, and nucleic acid which encode proteins yield essentially identical products. For example, due to the degeneracy of the genetic code, "silent substitutions" (*i.e.*, substitutions of a nucleic acid sequence which do not result in an alteration in an encoded protein) are an implied feature of every nucleic acid sequence which encodes an amino acid. As described herein, sequences are preferably optimized for expression in a particular host cell used to produce the chimeric glycosyltransferases (*e.g.*, yeast, human, and the like). Similarly, "conservative amino acid substitutions," in one or a few amino acids in an amino acid sequence are substituted with different amino acids with highly similar properties (*see*, the definitions section, *supra*), are also readily identified as being highly similar to a particular amino acid sequence, or to a particular nucleic acid sequence which encodes an amino acid. Such conservatively substituted variations of any particular sequence are a feature of the present invention. *See also*, Creighton (1984) *Proteins*, W.H. Freeman and Company. In addition, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are also "conservatively modified variations".

[0160] The practice of this invention can involve the construction of recombinant nucleic acids and the expression of genes in host cells, preferably bacterial host cells. Molecular cloning techniques to achieve these ends are known in the art. A wide variety of cloning and *in vitro* amplification methods suitable for the construction of recombinant nucleic acids such as expression vectors are well known to persons of skill. Examples of these techniques and instructions sufficient to direct persons of skill through many cloning exercises are found in Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology* volume 152 Academic Press, Inc., San Diego, CA (Berger); and *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1999 Supplement) (Ausubel). Suitable host cells for expression of the recombinant polypeptides are known to those of skill in the art, and include, for example, prokaryotic cells, such as *E. coli*, and eukaryotic cells including insect, mammalian and fungal cells (*e.g.*, *Aspergillus niger*)

[0161] Examples of protocols sufficient to direct persons of skill through *in vitro* amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Qβ-replicase amplification and other RNA polymerase mediated techniques are found in Berger, Sambrook, and Ausubel, as well as Mullis *et al.* (1987) U.S. Patent No. 4,683,202; *PCR Protocols A Guide to Methods and Applications* (Innis *et al.* eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (October 1, 1990) *C&EN* 36-47; *The Journal Of NIH Research* (1991) 3: 81-94; (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86: 1173; Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87: 1874; Lomell *et al.* (1989) *J. Clin. Chem.* 35: 1826; Landegren *et al.* (1988) *Science* 241: 1077-1080; Van Brunt (1990) *Biotechnology* 8: 291-294; Wu and Wallace (1989) *Gene* 4: 560; and Barringer *et al.* (1990) *Gene* 89: 117. Improved methods of cloning *in vitro* amplified nucleic acids are described in Wallace *et al.*, U.S. Pat. No. 5,426,039.

DETAILED DESCRIPTION OF THE INVENTION

I. Introduction

[0162] The present invention provides conditions for refolding eukaryotic glycosyltransferases that are expressed as insoluble proteins in bacterial inclusion bodies. Refolding buffers comprising redox couples are used to enhance refolding of insoluble eukaryotic glycosyltransferases. Refolding can also be enhanced by fusing a maltose binding domain to the insoluble eukaryotic glycosyltransferase. For some insoluble eukaryotic glycosyltransferases, refolding can also be enhanced by site directed mutagenesis to remove unpaired cysteines. Additional refolding enhancement can be provided be truncating a eukaryotic glycosyltransferase to remove, *e.g.*, a signal-anchor domain, a transmemebrane domain, and/or all or a portion of a stem region of the protein. The invention also provides methods to refold more than one glycosyltransferase in a single vessel, thereby enhancing refolding of the proteins and increasing efficiency of protein production. The refolded eukaryotic glycosyltransferases can be used to produce or to remodel polysaccharides, oligosaccharides, glycolipids, proteins, peptides, glycopeptides, and glycoproteins. The refolded eukaryotic glycosyltransferases can also be used to glycoPEGylate proteins, peptides, glycopeptides, or glycoproteins as described in PCT/US02/32263, which is herein incorporated by reference for all purposes.

II. Refolding insoluble glycosyltransferases

[0163] Many recombinant proteins expressed in bacteria are expressed as insoluble aggregates in bacterial inclusion bodies. Inclusion bodies are protein deposits found in both

the cytoplasmic and periplasmic space of bacteria. (*See, e.g.*, Clark, *Cur. Op. Biotech.* 12:202-207 (2001)). Eukaryotic glycosyltransferases are frequently expressed in bacterial inclusion bodies. Some eukaryotic glycosyltransferases are soluble in bacteria, *i.e.*, not produced in inclusion bodies, when only the catalytic domain of the protein is expressed. However, many eukaryotic glycosyltransferases remain insoluble and are expressed in bacterial inclusion bodies, even if only the catalytic domain is expressed, and methods for refolding these proteins to produce active glycosyltransferases are provided herein.

A. Conditions for refolding active glycosyltransferases

[0164] To produce active eukaryotic glycosyltranferases from bacterial cells, eukaryotic glycosyltranferases are expressed in bacterial inclusion bodies, the bacteria are harvested, disrupted and the inclusion bodies are isolated and washed. The proteins within the inclusion bodies are then solubilized. Solubilization can be performed using denaturants, *e.g.*, guanidinium chloride or urea; extremes of pH, such as acidic or alkaline conditions; or detergents.

[0165] After solubilization, denaturants are removed from the glycosyltransferase mixture. Denaturant removal can be done by a variety of methods, including dilution into a refolding buffer or buffer exchange methods. Buffer exchange methods include dialysis, diafiltration, gel filtration, and immobilization of the protein onto a solid support. (*See, e.g.*, Clark, *Cur. Op. Biotech.* 12:202-207 (2001)). Any of the above methods can be combined to remove denaturants.

[0166] Disulfide bond formation in the eukaryotic glycosyltransferase is promoted by addition of a refolding buffer comprising a redox couple. Redox couples include reduced and oxidized glutathione (GSH/GSSG), cysteine/cystine, cysteamine/cystamine, DTT/GSSG, and DTE/GSSG. (*See, e.g.*, Clark, *Cur. Op. Biotech.* 12:202-207 (2001), which is herein incorporated by reference for all purposes). In some embodiments, redox couples are added at an particular ratio of reduced to oxidized component, *e.g.*, 1/20, 20/1, ¼, 4/1, 1/10, 10/1, 1/2, 2/1, 1/5, 5/1, or 5/5.

[0167] Refolding can be performed in buffers at pH's ranging from, for example, 6.0 to 10.0. Refolding buffers can include other additives to enhance refolding, *e.g.*, L-arginine (0.4-1M); PEG; low concentrations of denaturants, such as urea (1-2M) and guanidinium chloride (0.5-1.5 M); and detergents (*e.g.*, Chaps, SDS, CTAB, lauryl maltoside, and Triton X-100).

[0168] Refolding can be over a given period of time, *e.g.*, for 1-48 hours, or overnight. Refolding can be done from about 4°C to about 40°C, including ambient temperatures.

[0169] A eukaryotic glycosyltransferase protein comprising a catalytic domain is expressed in bacterial inclusion bodies and then refolded using the above methods. Eukaryotic glycosyltransferases that comprise all or a portion of a stem region and a catalytic domain can also be used in the a methods described herein, as can eukaryotic glycosyltransferases comprising a catalytic domain fused to an MBP protein.

[0170] Those of skill will recognize that a protein has been refolded correctly when the refolded protein has detectable biological activity. For a glycosyltransferase biological activity is the ability to catalyze transfer of a donor substrate to an acceptor substrate, *e.g.*, a refolded ST3GalIII is able to transfer sialic acid to an acceptor substrate. Biological activity includes *e.g.*, specific activities of at least 1, 2, 5, 7, or 10 units of activity. Unit is defined as follows: one activity unit catalyzes the formation of 1 μmol of product per minute at a given temperature (*e.g.*, at 37°C) and pH value (*e.g.*, at pH 7.5). Thus, 10 units of an enzyme is a catalytic amount of that enzyme where 10 μmol of substrate are converted to 10 μmol of product in one minute at a temperature of, *e.g.*, 37 °C and a pH value of, *e.g.*, 7.5.

[0171] In one embodiment, eukaryotic ST3GalIII is expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0172] In one embodiment, eukaryotic GnT1 is expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0173] In one embodiment, eukaryotic GalT1 is expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0174] In one embodiment, eukaryotic St3GalI is expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0175] In one embodiment, eukaryotic St6 GalNAcTI is expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0176] In one embodiment, eukaryotic Core GalITI is expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0177] In one embodiment, eukaryotic GalNAcT2 is expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

B. Fusion of eukaryotic glycosyltransferases to maltose binding protein domains to enhance refolding

[0178] Maltose binding protein (MBP) domains are typically fused to proteins to enhance solubility of a the protein with a cell. *See, e.g.*, Kapust and Waugh *Pro. Sci.* 8:1668-1674 (1999). However, many eukaryotic glycosyltransferases, including truncated eukaryotic glycosyltransferases, remain insoluble when expressed in bacteria, even after fusion to a MBP domain. However, this application discloses that MBP domains can enhance refolding of insoluble eukaryotic glycosyltransferases after solubilization of the proteins from *e.g.*, an inclusion body. MBP domains from a variety of bacterial sources can be used in the invention, for example *Yersinia E. coli*, *Pyrococcus furiosus*, *Thermococcus litoralis*, *Thermatoga maritime*, and *Vibrio cholerae*, see, *e.g.*, Figure 43. In a preferred embodiment an *E. coli* MBP protein is fused to a eukaryotic glycosyltransferase. Amino acid linkers can be placed between the MBP domain and the glycosyltransferase. In another preferred embodiment, the MBP domain is fused to the amino terminus of the glycosyltransferase protein. The methods described above can be used to refold the MBP glycosyltransferase fusion proteins.

[0179] In one embodiment, a eukaryotic ST3GalIII protein is fused to an MBP domain, expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0180] In one embodiment, a eukaryotic GnT1 protein is fused to an MBP domain, expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0181] In one embodiment, a eukaryotic GalT1 protein is fused to an MBP domain, expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0182] In one embodiment, a eukaryotic St3GalI protein is fused to an MBP domain, expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0183] In one embodiment, a eukaryotic St6 GalNAcTI protein is fused to an MBP domain, expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0184] In one embodiment, a eukaryotic Core GalITI protein is fused to an MBP domain, expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0185] In one embodiment, a eukaryotic GalNAcT2 protein is fused to an MBP domain, expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0186] Additional amino acid tags can be added to an MBP-glycosyltransferase fusion. For example, purification tags can be added to enhance purification of the refolded protein. Purification tags include, *e.g.*, a polyhistidine tag, a glutathione S transferase (GST), a starch binding protein (SBP), an *E. coli* thioredoxin domain, a carboxy-terminal half of the SUMO protein, a FLAG epitope, and a myc epitope. Refolded glycosyltransferases can be further purified using a binding partner that binds to the purification tag. In a preferred embodiment, an MBP tag is fused to the eukaryotic glycosyltransferase to enhance refolding. Purification tags can be fused to MBP glycosyltransferase fusion protein including *e.g.*, GnT1, GalT1, StIII Gal3, St3GalI, St6 GalNAcTI, Core GalITI, or GalNAcT2.

[0187] In another embodiment, addition of an MBP domain to a protein can increase the expression of the protein. See, *e.g.*, example 12 where fusion of the SiaA protein to an MBP domain increased the expression of the protein. Other proteins with enhanced expression on fusion to MBP include *e.g.*, GnT1, GalT1, StIII Gal3, St3GalI, St6 GalNAcTI, Core GalITI, or GalNAcT2.

[0188] In another embodiment a self-cleaving protein tag, such as an intein, is included between the MPB domain and the glycosyltransferase to facilitate removal of the MBP domain after the fusion protein has been refolded. Inteins and kits for their use are commercially available, *e.g.*, from New England Biolabs.

C. Mutagenesis of glycosyltransferases to enhance refolding

[0189] Refolding of glycosyltransferases can also be enhanced by mutagenesis of the glycosyltransferase amino acid sequence. In one embodiment an unpaired cysteine residue is identified and mutated to enhance refolding of a glyscosyltransferase. In another embodiment, the amino terminus of the glycosyltransferase is truncated to remove a transmembrane domain, or to remove a transmembrane domain and all or a portion of the stem region of the protein. In a further embodiment, a glycosyltransferase is mutated to remove at least one unpaired cysteine residue and to truncate the amino terminus of the protein, *e.g.*, to remove a transmembrane domain, or to remove a transmembrane domain and all or a portion of the stem region of the protein. Once a glycosyltransferase nucleic acid sequence has been isolated, standard molecular biology methods can be used to change the nucleic acid sequence and thus the encoded amino acid sequence in a manner described herein.

1. Mutagenesis of unpaired cysteines in glycosyltransferases to enhance refolding

[0190] As refolding occurs, cysteine residues in a denatured protein form disulfide bonds that help to reproduce the structure of the active protein. Incorrect pairing of cysteine residues can lead to protein misfolding. Proteins with unpaired cysteine residues are susceptible to misfolding because a normally unpaired cysteine can form a disulfide bond with normally paired cysteine making correct cysteine pairing and protein refolding impossible. Thus, one method to enhance refolding of a particular glycosyltransferase is to identify unpaired cysteine residues and remove them.

[0191] Unpaired cysteine residues can be identified by determining the structure of the glycosyltransferase of interest. Protein structure can be determined based on actual data for the glycosyltransferase of interest, *e.g.*, circular dichroism, NMR, and X-ray crystallography. Protein structure can also be determined using computer modeling. Computer modeling is a technique that can be used to model related structures based on known three-dimensional structures of homologous molecules. Standard software is commercially available. (*See e.g.*, www.accelrys.com for the multitude of software available to do computer modeling.) Once an unpaired cysteine residue is identified, the DNA encoding the glycosyltransferase of interest can be mutated using standard molecular biology techniques to remove the unpaired cysteine, by deletion or by substitution with another amino acid residue. Computer modeling is used again to select an amino acid of appropriate size, shape, and charge for substitution.

Unpaired cysteines can also be determined by peptide mapping. Once the glycosyltransferase of interest is mutated, the protein is expressed in bacterial inclusion bodies and refolding ability is determined. A correctly refolded glycosyltransferase will have biological activity.

[0192] In preferred embodiments, the following amino acid residues are substituted for an unpaired cysteine residue in a eukaryotic glycosyltransferase to enhance refolding: Ala, Ser, Thr, Asp, Ile, or Val. Gly can also be used if the unpaired cysteine is not in a helical structure.

[0193] Human N-acetylglucosaminyltransferase I (GnTI, accession number NP_002397) is an example of a glycosyltransferase that exhibited enhanced refolding after mutagenesis of an unpaired cysteine. (*See, e.g.*, Example 2, below.) Human GNTI is closely related to a number of eukaryotic GNTI proteins, *e.g.*, Chinese hamster, accession number AAK61868; rabbit accession number AAA31493; rat accession number NP_110488; golden hamster, accession number AAD04130; mouse, accession number P27808; zebrafish, accession number AAH58297; *Xenopus*, accession number CAC51119; *Drosophila*, accession number NP_525117; *Anopheles*, accession number XP_315359; *C. elegans*, accession number NP_497719; *Physcomitrella patens*, accession number CAD22107; *Solanum tuberosum*, accession number CAC80697; *Nicotiana tabacum*, accession number CAC80702; *Oryza sativa*, accession number CAD30022; *Nicotiana benthamiana*, accession number CAC82507; and *Arabidopsis thaliana*, accession number NP_195537.

[0194] The structure of the rabbit *N*-acetylglucosaminyltransferase I (GnTI) protein had been determined and showed that CYS123 was unpaired. (Amino acid residue numbers refer to the full length protein sequence even when a GNTI protein has been truncated.) Computer modeling based on the rabbit GnTI was used to determine the structure of the human GnTI protein. An alignment is shown in Figure 6. In the human GnTI protein, CYS121 was unpaired. Substitutions for CYS121 were made in human GnTI. A CYS121SER mutant and a CYS121ALA mutant were active. In contrast, a CYS121THR mutant had no detectable activity and a CYS121ASP mutant had low activity. A double mutant, ARG120ALA, CYS121HIS, was constructed based on the predicted structure of the *C. elegans* GNT1 protein, and had activity.

[0195] The amino acid sequences of the eukaryotic GnTI proteins listed above can be used to determine protein structure based on computer modeling and the conserved function of CYS123 from rabbit and CYS121 from human. Based on that analysis, residue 123 is an

unpaired cysteine in the following proteins: Chinese hamster GnTI, the rabbit GnTI, the rat GnTI, the golden hamster GnTI, and the mouse GnTI. Thus, CYS123 can be mutated in each of the GnTI enzymes to serine, alanine, or arginine to produce an active protein with enhanced refolding activity. The following double mutant in the above proteins, ARG122ALA CYS123HIS, will also exhibit enhanced refolding.

[0196] In one embodiment, any of the eukaryotic GnT1 proteins listed above is mutated to remove an unpaired cysteine residue, *e.g.*, CYS121SER, CYS121ALA, CYS121ASP, or the double mutant ARG120ALA CYS121HIS, expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0197] Another glycosyltransferase that exhibits enhanced refolding on mutation of an unpaired cysteines is Gal T1. Cysteine 342 was mutated to a threonine residue in the bovine Gal T1 and the mutated enzyme exhibited enhanced refolding after solubilization. *See, e.g.*, Ramakrishnan *et al. J. Biol. Chem.* 276:37666-37671 (2001). Of interest, mutation of an unpaired cysteine to threonine in the GnTI enzyme, abolished activity.

[0198] In one embodiment, a Gal T1 protein is mutated to remove an unpaired cysteine residue, *e.g.*, CYS342THR, expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0199] Another glycosyltransferase that exhibits enhanced refolding on mutation of an unpaired cysteines is GalNAc T2. Many amino acids residues are shared between the GalNAc T2 protein and the GalNAc T1 protein. After mutation of two cysteines residues, CYS212 and CYS214, the human GalNAc T1 protein remained active when expressed from COS cells. *See, e.g.*, Tenno *et al.*, *Eur. J. Biochem.* 269:4308-4316 (2002). The active mutations included CYS212ALA, CYS214ALA, CYS212SER, CYS214SER, and a double mutant CYS212SER CYS214SER. Cysteine residues corresponding to human GalNAc T1 CYS212 and CYS214 residues are conserved in the human GalNAc T2 protein, *i.e.*, residues CYS227 and CYS229. *See, e.g.*, Figure 44. Thus, a GalNAc T2 protein comprising one of the following mutations can be used to enhance refolding of the insoluble protein: CYS227ALA, CYS229ALA, CYS227SER, CYS229SER, and a double mutant CYS227SER CYS229SER. The numbering of residues refers to human GalNAc T2 proteins, but conserved cysteines residues that correspond to CYS227 and CYS229 can be identified in

other eukaryotic GalNAc T2 proteins, *e.g.*, mouse, rat, rabbit, pig, and mutated to enhance refolding.

[0200] In one embodiment, a GalNAc T2 protein is mutated to remove an unpaired cysteine residue, *e.g.*, CYS227ALA, CYS229ALA, CYS227SER, CYS229SER, or a double mutant CYS227SER CYS229SER, expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

[0201] Another glycosyltransferase that exhibits enhanced refolding on mutation of an unpaired cysteines is Core 1 Gal T1. In one embodiment, the drosophila Core 1 Gal T1 is mutated. In another embodiment, the human Core 1 Gal T1 is mutated. The drosophila Core 1 Gal T1 protein has seven cysteine residues. Each cysteine residue is mutated individually to either serine or alanine. The mutated drosophila Core 1 Gal T1 proteins are expressed in *E. coli* inclusion bodies, solubilized, and refolded. Enzymatic activity of the refolded mutant drosophila Core 1 Gal T1 is assayed and compared to wild type refolded drosophila Core 1 Gal T1 activity. Enhanced refolding is indicated by an increase in enzymatic activity in a mutant drosophila Core 1 Gal T1 as compared to the wild type protein.

[0202] In one embodiment, a Core 1 Gal T1 protein is mutated to remove an unpaired cysteine residue, *e.g.*, either the drosophila protein or the human protein, expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine. Preferred cysteine residues for substitution in the drosophila Core 1 Gal T1 are C103, C127, C208, C246, C261, C315 and C316.

2. Truncation of glycosyltransferases to enhance refolding

[0203] Eukaryotic glycosyltransferases generally include the following domains: a catalytic domain, a stem region, a transmembrane domain, and a signal-anchor domain. When expressed in bacteria, the signal anchor domain, and transmembrane domains are typically deleted. Eukaryotic glycosyltransferases used in the methods of the invention can include all or a portion of the stem region and the catalytic domain. In some embodiments, the eukaryotic glycosyltransferases comprise only the catalytic domain.

[0204] Glycosyltransferase domains can be identified for deletion mutagenesis. For example, those of skill in the art can identify a stem region in a eukaryotic glycosyltransferase and delete stem region amino acids one by one to identify truncated eukaryotic glycosyltransferase proteins with high activity on refolding.

[0205] The deletion mutants in this application are referenced in two ways: Δ or D followed by the number of residues deleted from the amino terminus of the native full length amino acid sequence, or by the symbol and residue number of the first amino acid residue translated from the native full length amino acid sequence. For example, ST6GalNAcI Δ or D35 and ST6GalNAcI K36, both refer to the same truncation of the human ST6GalNAcI protein.

[0206] For example, the rat ST3GalIII protein includes a stem region from about amino acid residues 29-84. The catalytic domain of the protein comprises amino acids from about residue 85-374. Thus, a truncated rat ST3GalIII protein can have deletions at the amino terminus of about *e.g.*, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, or 85 residues.

[0207] Deletion mutations can also be made in a GnT1 protein. For example, the human GnT1 protein includes a stem region from about amino acid residues 31-112. Thus, a truncated human GnT1 protein can have deletions at the amino terminus of about *e.g.*, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, or 111 residues.

[0208] Deletion mutations can also be made in a Gal T1 protein. For example, the bovine GalT1 protein includes a stem region from about amino acid residues 71-129. Thus, a truncated bovine GalT1 protein can have deletions at the amino terminus of about *e.g.*, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97 98, 99, 100, 101, 102, 103, 104, 105, 106, 017, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, or 128 residues.

[0209] Deletion mutations can also be made in a Core1 GalT1 protein. For example, the Drosophila Core1 GalT1 protein includes a stem region from about amino acid residues 36-102. Thus, a truncated Drosophila Core1 GalT1 protein can have deletions at the amino terminus of about *e.g.*, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97 98, 99, 100, 101, or 102 residues. As another example, the human Core1 GalT1 protein includes a stem region from

about amino acid residues 32-90. Thus, a truncated human Core1 GalT1 protein can have deletions at the amino terminus of about *e.g.*, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90 residues.

[0210] Deletion mutations can also be made in an ST3Gal1 protein. For example, the human ST3Gal1 protein includes a stem region from about amino acid residues 18-58. Thus, a truncated human ST3Gal1 protein can have deletions at the amino terminus of about *e.g.*, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, or 58 residues. As another example, the porcine ST3Gal1 protein includes a stem region from about amino acid residues 28-61. Thus, a truncated porcine ST3Gal1 protein can have deletions at the amino terminus of about *e.g.*, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, or 61 residues.

[0211] Deletion mutations can also be made in a GalNAcT2 protein. For example, the rat GalNAcT2 protein includes a stem region from about amino acid residues 40-95. Thus, a truncated rat GalNAcT2 protein can have deletions at the amino terminus of about *e.g.*, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, or 95 residues.

[0212] Deletion mutations can also be made in an ST6GalNAcI protein. For example, the mouse ST6GalNAcI protein includes a stem region from about amino acid residues 30-207. Thus, a truncated mouse ST6GalNAcI protein can have deletions at the amino terminus of about *e.g.*, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97 98, 99, 100, 101, 102, 103, 104, 105, 106, 017, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197 198, 199, 200, 201, 202, 203, 204, 205, 206, or 207 residues. As another example, the human ST6GalNAcI protein includes a stem region from about amino

acid residues 35-278. Thus, a truncated human ST6GalNAcI protein can have deletions at the amino terminus of about *e.g.*, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97 98, 99, 100, 101, 102, 103, 104, 105, 106, 017, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277,or 278 residues. As still another example, chicken ST6GalNAcI protein includes a stem region from about amino acid residues 37-253. Thus, a truncated chicken ST6GalNAcI protein can have deletions at the amino terminus of about *e.g.*, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97 98, 99, 100, 101, 102, 103, 104, 105, 106, 017, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, or 253 residues.

D. One pot refolding of glycosyltransferases

[0213] These embodiments of the invention are based on the surprising observation that multiple eukaryotic glycosyltransferases expressed in bacterial inclusion bodies can be refolded in a single vessel, *i.e.*, a one pot method. Using this method at least two glycosyltransferases can be refolded together resulting in savings of time and materials.

Refolding conditions are described above. The refolding conditions are optimized for the mixture of glycosyltransferases, thus, conditions may not be optimal for any particular enzyme in the mixture. However, because refolding is optimized for the combination of glycosyltransferases, each of the refolded glycosyltransferases in the end product has detectable biological activity. Biological activity refers to enzymatic activity of the refolded enzymes and can be expressed as specific activity. Biological activity includes *e.g.*, specific activities of at least 0.1, 0.5, 1, 2, 5, 7, or 10 units of activity. Unit is defined as follows: one activity unit catalyzes the formation of 1 μmol of product per minute at a given temperature (*e.g.*, at 37°C) and pH value (*e.g.*, at pH 7.5). Thus, 10 units of an enzyme is a catalytic amount of that enzyme where 10 μmol of substrate are converted to 10 μmol of product in one minute at a temperature of, *e.g.*, 37 °C and a pH value of, *e.g.*, 7.5. The reaction mixture comprising refolded glycosyltransferases can then be used *e.g.*, to synthesize oligosaccharides, to synthesize glycolipids, to remodel glycoproteins, and to glycoPEGlyate glycoproteins.

[0214] In some embodiments, the glycosyltransferases can be solubilized individually from inclusion bodies and then combined under conditions appropriate for refolding. In other embodiments, inclusion bodies containing glycosyltransferases are combined, solubilized, and then refolded under appropriate conditions.

[0215] Refolding buffers typically include a redox couple. Refolding can be performed at pH's ranging from, for example, 6.0 to 10.0. Refolding buffers can include other additives to enhance refolding, *e.g.*, L-arginine (0.4-1M); PEG; low concentrations of denaturants, such as urea (1-2M) and guanidinium chloride (0.5-1.5 M); and detergents (*e.g.*, Chaps, SDS, CTAB, and Triton X-100).

[0216] In some embodiments, refolding is performed in a stationary vessel, *i.e.*, without mixing, stirring, shaking or otherwise moving the reaction mixture.

[0217] The combination of refolded enzymes can include enzymes to construct a particular oligosaccharide structure. Those of skill will be able to identify appropriate glycosyltransferases for inclusion in the mixture once a desired end product is identified.

[0218] The reaction mixtures of refolded enzymes can include glycosyltransferases that have been mutated to enhance refolding, *e.g.*, the GnTI enzymes described above.

[0219] In a preferred embodiment, enzymes that perform N-linked glycosylation steps are refolded together in a single vessel. For example, *N*-acetylglucosaminyltransferase I (GnTI), β-1,4 galactosyltransferase I (Gal TI), and *N*-acetyllactosaminide α-2,3-sialyltransferase (ST3GalIII) can be expressed in bacterial inclusion bodies, solubilized, and refolded together in a single vessel. The end product exhibited activity of all three proteins, indicating they were all correctly refolded. Refolding also occurred when GnTI and Gal TI were refolded together without ST3GalIII. The experiments are described in detail in Example 3.

[0220] In another preferred embodiment, O-linked glycosylation of a peptide or protein is accomplished using the bacterially expressed and refolded glycosyltransferases of this disclosure. For example, a refolded MBP-GalNAcT2(D51) enzyme can be used to add GalNAc to polypeptides. *E.g.*, example 4 provides a demonstration that refolded MBP-GalNAcT2(D51) can be used to add GalNAc to the GCSF protein. Combinations of O-linked glycosyltransferases can be used to remodel *e.g.*, proteins, peptides, glycoproteins or glycopeptides. Those combinations include *e.g.*, GalNAc-T2 and ST6GalNAc1; or GalNAc-T2, core 1 GalT1 and ST3Gal1 or ST3GalT2

III. Glycosyltransferases

[0221] The glycosyltransferases of use in practicing the present invention are eukaryotic glycosyltransferases. Examples of such glycosyltransferases include those described in Staudacher, E. (1996) *Trends in Glycoscience and Glycotechnology*, **8**: 391-408, afmb.cnrs-mrs.fr/~pedro/CAZY/gtf.html and www.vei.co.uk/TGN/gt_guide.htm, but are not limited thereto.

Eukaryotic glycosyltransferases

[0222] Some eukaryotic glycosyltransferases have topological domains at their amino terminus that are not required for catalytic activity (*see,* US Patent No. 5, 032,519). Of the glycosyltransferases characterized to date, the "cytoplasmic domain," is most commonly between about 1 and about 10 amino acids in length, and is the most amino-terminal domain; the adjacent domain, termed the "signal-anchor domain," is generally between about 10-26 amino acids in length; adjacent to the signal-anchor domain is a "stem region," which is generally between about 20 and about 60 amino acids in length, and known to function as a retention signal to maintain the glycosyltransferase in the Golgi apparatus; and at the carboxyl side of the stem region is the catalytic domain.

[0223] Many mammalian glycosyltransferases have been cloned and expressed and the recombinant proteins have been characterized in terms of donor and acceptor substrate specificity and they have also been investigated through site directed mutagenesis in attempts to define residues or domains involved in either donor or acceptor substrate specificity (Aoki *et al.* (1990) *EMBO. J.* 9: 3171-3178; Harduin-Lepers *et al.* (1995) *Glycobiology* 5(8): 741-758; Natsuka and Lowe (1994) *Current Opinion in Structural Biology* 4: 683-691; Zu *et al.* (1995) *Biochem. Biophys. Res. Comm.* 206(1): 362-369; Seto *et al.* (1995) *Eur. J. Biochem.* 234: 323-328; Seto *et al.* (1997) *J. Biol. Chem.* 272: 14133-141388).

[0224] In one group of embodiments, a functional domain of the recombinant glycosyltransferase proteins of the present invention is obtained from a known sialyltransferase. Examples of sialyltransferases that are suitable for use in the present invention include, but are not limited to, ST3GalIII, ST3Gal IV, ST3Gal I, ST6Gal I, ST3Gal V, ST6Gal II, ST6GalNAc I, ST6GalNAc II, and ST6GalNAc III (the sialyltransferase nomenclature used herein is as described in Tsuji *et al.* (1996) *Glycobiology* 6: v-xiv). An exemplary α2,3-sialyltransferase (EC 2.4.99.6) transfers sialic acid to the non-reducing terminal Gal of a Galβ1→4GlcNAc disaccharide or glycoside. *See*, Van den Eijnden *et al., J. Biol. Chem.*, 256:3159 (1981), Weinstein *et al., J. Biol. Chem.*, 257:13845 (1982) and Wen *et al., J. Biol. Chem.*, 267:21011 (1992). Another exemplary α2,3-sialyltransferase (EC 2.4.99.4) transfers sialic acid to the non-reducing terminal Gal of a Galβ1→3GalNAc disaccharide or glycoside. *See*, Rearick *et al., J. Biol. Chem.*, 254: 4444 (1979) and Gillespie *et al., J. Biol. Chem.*, 267:21004 (1992). Further exemplary enzymes include Gal-β-1,4-GlcNAc α-2,6 sialyltransferase (*See*, Kurosawa *et al. Eur. J. Biochem.* 219: 375-381 (1994)). Sialyltransferase nomenclature is described in Tsuji, S. *et al.* (1996) *Glycobiology* 6:v-vii.

[0225] An example of a sialyltransferase that is useful in the claimed methods is ST3GalIII, which is also referred to as α(2,3)sialyltransferase (EC 2.4.99.6). This enzyme catalyzes the transfer of sialic acid to the Gal of a Galβ1,3GlcNAc, Galβ1,3GalNAc or Galβ1,4GlcNAc glycoside (*see, e.g.*, Wen *et al.* (1992) *J. Biol. Chem.* **267**: 21011; Van den Eijnden *et al.* (1991) *J. Biol. Chem.* **256:** 3159). The sialic acid is linked to a Gal with the formation of an α-linkage between the two saccharides. Bonding (linkage) between the saccharides is between the 2-position of NeuAc and the 3-position of Gal. This particular enzyme can be isolated from rat liver (Weinstein *et al.* (1982) *J. Biol. Chem.* **257**: 13845); the human cDNA (Sasaki *et al.* (1993) *J. Biol. Chem.* **268**: 22782-22787; Kitagawa & Paulson (1994) *J. Biol.*

Chem. **269**: 1394-1401) and genomic (Kitagawa *et al.* (1996) *J. Biol. Chem.* **271**: 931-938) DNA sequences are known, facilitating production of this enzyme by recombinant expression. In a preferred embodiment, the claimed sialylation methods use a rat ST3GalIII. Rat ST3GalIII has been cloned and and the sequence is known. *See, e.g.*, Wen *et al.*, *J. Biol. Chem.* 267:21011-21019 (1992) and Accession number M97754.

[0226] In another group of embodiments, a functional domain of the recombinant glycosyltransferase proteins of the present inventions is obtained from a fucosyltransferase. A number of fucosyltransferases are known to those of skill in the art. Briefly, fucosyltransferases include any of those enzymes which transfer L-fucose from GDP-fucose to a hydroxy position of an acceptor sugar. In some embodiments, for example, the acceptor sugar is a GlcNAc in a Galβ(1→4)GlcNAc group in an oligosaccharide glycoside. Suitable fucosyltransferases for this reaction include the known Galβ (1→3,4)GlcNAc α(1→3,4)fucosyltransferase (FTIII, E.C. No. 2.4.1.65) which is obtained from human milk (*see*, Palcic, *et al., Carbohydrate Res.* 190:1-11 (1989); Prieels, *et al., J. Biol. Chem.* 256: 10456-10463 (1981); and Nunez, *et al., Can. J. Chem.* 59: 2086-2095 (1981)) and the Galβ(1→4)GlcNAc α(1→3)fucosyltransferases (FTIV, FTV, and FTVI, E.C. No. 2.4.1.65) and NeuAcα(2,3)βGal(1→4)βGlcNAc α(1→3)fucosyltransferases (FTVII) which are found in human serum. Also, available is the α1,3 fucosyltransferase IX (nucleotide sequences of human and mouse FTIX) as described in Kaneko *et al.* (1999) *FEBS Lett.* **452**: 237-242. In addition, a recombinant form of Galβ (1→3,4)GlcNAc α(1→3,4)fucosyltransferase is available (*see*, Dumas, *et al., Bioorg. Med. Letters* 1:425-428 (1991) and Kukowska-Latallo, *et al., Genes and Development* 4:1288-1303 (1990)). Other exemplary fucosyltransferases include α1,2 fucosyltransferase (E.C. No. 2.4.1.69). Enzymatic fucosylation can be carried out by the methods described in Mollicone, *et al., Eur. J. Biochem.* 191:169-176 (1990) or U.S. Patent No. 5,374,655.

[0227] In another group of embodiments, a functional domain of the recombinant glycosyltransferase proteins of the present inventions is obtained from known galactosyltransferases. Exemplary galactosyltransferases include β-1,4 galactosyltransferase I, α1,3- galactosyltransferases (E.C. No. 2.4.1.151, *see, e.g.*, Dabkowski *et al., Transplant Proc.* **25**:2921 (1993) and Yamamoto *et al. Nature* 345:229-233 (1990), bovine (GenBank j04989, Joziasse *et al.* (1989) *J. Biol. Chem.* 264:14290-14297), murine (GenBank m26925; Larsen *et al.* (1989) *Proc. Nat'l. Acad. Sci. USA* 86:8227-8231), porcine (GenBank L36152;

Strahan *et al* (1995) *Immunogenetics* 41:101-105)). Another suitable α1,3-galactosyltransferase is that which is involved in synthesis of the blood group B antigen (EC 2.4.1.37, Yamamoto *et al.* (1990) *J. Biol. Chem.* 265:1146-1151 (human)). Also suitable for use in the fusion proteins of the invention are α1,4-galactosyltransferases, which include, for example, EC 2.4.1.90 (LacNAc synthetase) and EC 2.4.1.22 (lactose synthetase) (bovine (D'Agostaro et al (1989) Eur. J. Biochem. 183:211-217), human (Masri *et al.* (1988) *Biochem. Biophys. Res. Commun.* 157:657-663), murine (Nakazawa et al (1988) *J. Biochem.* 104:165-168), as well as E.C. 2.4.1.38 and the ceramide galactosyltransferase (EC 2.4.1.45, Stahl *et al.* (1994) *J. Neurosci. Res.* 38:234-242). Other suitable galactosyltransferases include, for example, α1,2-galactosyltransferases (from *e.g.*, *Schizosaccharomyces pombe*, Chapell *et al* (1994) *Mol. Biol. Cell* 5:519-528).

[0228] Other glycosyltransferases that are useful in the recombinant fusion proteins of the present invention have been described in detail, as for the sialyltransferases, galactosyltransferases, and fucosyltransferases. In particular, the glycosyltransferase can also be, for instance, a glucosyltransferase, *e.g.*, Alg8 (Stagljov *et al., Proc. Natl. Acad. Sci. USA* 91:5977 (1994)) or Alg5 (Heesen *et al. Eur. J. Biochem.* 224:71 (1994)), N-acetylgalactosaminyltransferases such as, for example, β(1,3)-*N*-acetylgalactosaminyltransferase, β(1,4)-*N*-acetylgalactosaminyltransferases (US Patent No. 5,691,180, Nagata *et al. J. Biol. Chem.* 267:12082-12089 (1992), and Smith *et al. J. Biol Chem.* 269:15162 (1994)) and protein *N*-acetylgalactosaminyltransferase (Homa *et al. J. Biol Chem.* 268:12609 (1993)). Suitable *N*-acetylglucosaminyltransferases include GnTI (2.4.1.101, Hull *et al., BBRC* 176:608 (1991)), GnTII, and GnTIII (Ihara *et al. J. Biochem.* 113:692 (1993)), GnTV (Shoreiban *et al. J. Biol. Chem.* 268: 15381 (1993)), O-linked *N*-acetylglucosaminyltransferase (Bierhuizen *et al. Proc. Natl. Acad. Sci. USA* **89:**9326 (1992)), N-acetylglucosamine-1-phosphate transferase (Rajput *et al. Biochem J.***285:**985 (1992), and hyaluronan synthase. Also of interest are enzymes involved in proteoglycan synthesis, such as, for example, *N*-acetylgalactosaminyltransferase I (EC 2.4.1.174), and enzymes involved in chondroitin sulfate synthesis, such as *N*-acetylgalactosaminyltransferase II (EC 2.4.1.175). Suitable mannosyltransferases include α(1,2) mannosyltransferase, α(1,3) mannosyltransferase, β(1,4) mannosyltransferase, Dol-P-Man synthase, OCh1, and Pmt1. Xylosyltransferases include, for example, protein xylosyltransferase (EC 2.4.2.26).

[0229] In some embodiments, eukaryotic *N*-acetylgalactosaminyltransferases are expressed in bacteria and refolded using the methods of this disclosure. A number of GalNAcT enzymes have been isolated and characterized, *e.g.*, GalNAcT1, accession number X85018; GalNAcT2, accession number X85019 (both described in White *et al.*, J. Biol. Chem. 270:24156-24165 (1995)); and GalNAcT3, accession number X92689 (described in Bennett *et al.*, J. Biol. Chem. 271:17006-17012 (1996)).

IV. Nucleic acids

[0230] Nucleic acids that encode glycosyltransferases, and methods of obtaining such nucleic acids, are known to those of skill in the art. Suitable nucleic acids (*e.g.*, cDNA, genomic, or subsequences (probes)) can be cloned, or amplified by *in vitro* methods such as the polymerase chain reaction (PCR), the ligase chain reaction (LCR), the transcription-based amplification system (TAS), or the self-sustained sequence replication system (SSR). A wide variety of cloning and *in vitro* amplification methodologies are well-known to persons of skill. Examples of these techniques and instructions sufficient to direct persons of skill through many cloning exercises are found in Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology* 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.* (1989) *Molecular Cloning - A Laboratory Manual* (2nd ed.) Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY, (Sambrook *et al.*); *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1994 Supplement) (Ausubel); Cashion *et al.*, U.S. patent number 5,017,478; and Carr, European Patent No. 0,246,864.

[0231] A DNA that encodes a glycosyltransferase, or a subsequences thereof, can be prepared by any suitable method described above, including, for example, cloning and restriction of appropriate sequences with restriction enzymes. In one preferred embodiment, nucleic acids encoding glycosyltransferases are isolated by routine cloning methods. A nucleotide sequence of a glycosyltransferase as provided in, for example, GenBank or other sequence database (see above) can be used to provide probes that specifically hybridize to a glycosyltransferase gene in a genomic DNA sample, or to an mRNA, encoding a glucosyltransferase, in a total RNA sample (*e.g.*, in a Southern or Northern blot). Once the target nucleic acid encoding a glycosyltransferase is identified, it can be isolated according to standard methods known to those of skill in the art (*see, e.g.*, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual, 2nd Ed., Vols. 1-3*, Cold Spring Harbor

Laboratory; Berger and Kimmel (1987) *Methods in Enzymology, Vol. 152: Guide to Molecular Cloning Techniques,* San Diego: Academic Press, Inc.; or Ausubel *et al.* (1987) *Current Protocols in Molecular Biology,* Greene Publishing and Wiley-Interscience, New York). Further, the isolated nucleic acids can be cleaved with restriction enzymes to create nucleic acids encoding the full-length glycosyltransferse, or subsequences thereof, *e.g.*, containing subsequences encoding at least a subsequence of a stem region or catalytic domain of a glycosyltransferase. These restriction enzyme fragments, encoding a glycosyltransferase or subsequences thereof, may then be ligated, for example, to produce a nucleic acid encoding a recombinant glycosyltransferase fusion protein.

[0232] A nucleic acid encoding a glycosyltransferase, or a subsequence thereof, can be characterized by assaying for the expressed product. Assays based on the detection of the physical, chemical, or immunological properties of the expressed protein can be used. For example, one can identify a cloned glycosyltransferase, including a glycosyltransferase fusion protein, by the ability of a protein encoded by the nucleic acid to catalyze the transfer of a saccharide from a donor substrate to an acceptor substrate. In a preferred method, capillary electrophoresis is employed to detect the reaction products. This highly sensitive assay involves using either saccharide or disaccharide aminophenyl derivatives which are labeled with fluorescein as described in Wakarchuk *et al.* (1996) *J. Biol. Chem.* 271 (45): 28271-276. For example, to assay for a *Neisseria lgtC* enzyme, either FCHASE-AP-Lac or FCHASE-AP-Gal can be used, whereas for the *Neisseria lgtB* enzyme an appropriate reagent is FCHASE-AP-GlcNAc (*Id.*).

[0233] Also, a nucleic acid encoding a glycosyltransferase, or a subsequence thereof, can be chemically synthesized. Suitable methods include the phosphotriester method of Narang *et al.* (1979) *Meth. Enzymol.* 68: 90-99; the phosphodiester method of Brown *et al.* (1979) *Meth. Enzymol.* 68: 109-151; the diethylphosphoramidite method of Beaucage *et al.* (1981) *Tetra. Lett.*, 22: 1859-1862; and the solid support method of U.S. Patent No. 4,458,066. Chemical synthesis produces a single stranded oligonucleotide. This can be converted into double stranded DNA by hybridization with a complementary sequence, or by polymerization with a DNA polymerase using the single strand as a template. One of skill recognizes that while chemical synthesis of DNA is often limited to sequences of about 100 bases, longer sequences may be obtained by the ligation of shorter sequences.

[0234] Nucleic acids encoding glycosyltransferases, or subsequences thereof, can be cloned using DNA amplification methods such as polymerase chain reaction (PCR). Thus, for example, the nucleic acid sequence or subsequence is PCR amplified, using a sense primer containing one restriction enzyme site (*e.g., Nde*I) and an antisense primer containing another restriction enzyme site (*e.g., Hin*dIII). This will produce a nucleic acid encoding the desired glycosyltransferase or subsequence and having terminal restriction enzyme sites. This nucleic acid can then be easily ligated into a vector containing a nucleic acid encoding the second molecule and having the appropriate corresponding restriction enzyme sites. Suitable PCR primers can be determined by one of skill in the art using the sequence information provided in GenBank or other sources. Appropriate restriction enzyme sites can also be added to the nucleic acid encoding the glycosyltransferase protein or protein subsequence by site-directed mutagenesis. The plasmid containing the glycosyltransferase-encoding nucleotide sequence or subsequence is cleaved with the appropriate restriction endonuclease and then ligated into an appropriate vector for amplification and/or expression according to standard methods. Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods are found in Berger, Sambrook, and Ausubel, as well as Mullis *et al.*, (1987) U.S. Patent No. 4,683,202; *PCR Protocols A Guide to Methods and Applications* (Innis *et al.*, eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (October 1, 1990) *C&EN* 36-47; *The Journal Of NIH Research* (1991) 3: 81-94; (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86: 1173; Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87, 1874; Lomell *et al.* (1989) *J. Clin. Chem.*, 35: 1826; Landegren *et al.*, (1988) *Science* 241: 1077-1080; Van Brunt (1990) *Biotechnology* 8: 291-294; Wu and Wallace (1989) *Gene* 4: 560; and Barringer *et al.* (1990) *Gene* 89: 117.

[0235] Other physical properties of a cloned glycosyltransferase protein, including glycosyltransferase fusion protein, expressed from a particular nucleic acid, can be compared to properties of known glycosyltransferases to provide another method of identifying suitable sequences or domains of the glycosyltransferase that are determinants of acceptor substrate specificity and/or catalytic activity. Alternatively, a putative glycosyltransferase gene or recombinant glycosyltransferase gene can be mutated, and its role as glycosyltransferase, its ability to be refolded, or the role of particular sequences or domains established by detecting a variation in the structure of a carbohydrate normally produced by the unmutated, naturally-occurring, or control glycosyltransferase.

[0236] Functional domains of cloned glycosyltransferases can be identified by using standard methods for mutating or modifying the glycosyltransferases and testing the modified or mutated proteins for activities such as acceptor substrate activity and/or catalytic activity, as described herein. The functional domains of the various glycosyltransferases can be used to construct nucleic acids encoding recombinant glycosyltransferase fusion proteins comprising the functional domains of one or more glycosyltransferases. These fusion proteins can then be tested for the desired acceptor substrate or catalytic activity.

[0237] In an exemplary approach to cloning recombinant glycosyltransferase fusion proteins, the known nucleic acid or amino acid sequences of cloned glycosyltransferases are aligned and compared to determine the amount of sequence identity between various glycosyltransferases. This information can be used to identify and select protein domains that confer or modulate glycosyltransferase activities, *e.g.*, acceptor substrate activity and/or catalytic activity based on the amount of sequence identity between the glycosyltransferases of interest. For example, domains having sequence identity between the glycosyltransferases of interest, and that are associated with a known activity, can be used to construct recombinant glycosyltransferase fusion proteins containing that domain, and having the activity associated with that domain (*e.g.*, acceptor substrate specificity and/or catalytic activity).

V. Expression of recombinant glycosyltransferases

[0238] Recombinant eukaryotic glycosyltransferases can be expressed in a variety of host cells, including *E. coli,* other bacterial hosts, yeast, and various higher eukaryotic cells such as the COS, CHO and HeLa cells lines and myeloma cell lines. The host cells can be mammalian cells, plant cells, or microorganisms, such as, for example, yeast cells, bacterial cells, or filamentous fungal cells. Examples of suitable host cells include, for example, *Azotobacter sp.* (*e.g.*, *A. vinelandii*), *Pseudomonas sp.*, *Rhizobium sp.*, *Erwinia sp.*, *Escherichia sp.* (*e.g.*, *E. coli*), *Bacillus, Pseudomonas, Proteus, Salmonella, Serratia, Shigella, Rhizobia, Vitreoscilla, Paracoccus* and *Klebsiella sp.*, among many others. The cells can be of any of several genera, including Saccharomyces (*e.g.*, *S. cerevisiae*), Candida (*e.g.*, *C. utilis, C. parapsilosis, C. krusei, C. versatilis, C. lipolytica, C. zeylanoides, C. guilliermondii, C. albicans, and C. humicola*), Pichia (*e.g.*, *P. farinosa and P. ohmeri*), Torulopsis (*e.g.*, *T. candida, T. sphaerica, T. xylinus, T. famata, and T. versatilis*), Debaryomyces (*e.g.*, *D. subglobosus, D. cantarellii, D. globosus, D. hansenii, and D. japonicus*), Zygosaccharomyces (*e.g.*, *Z. rouxii and Z. bailii*), Kluyveromyces (*e.g.*, *K.*

marxianus), Hansenula (*e.g., H. anomala and H. jadinii*), and Brettanomyces (*e.g., B. lambicus and B. anomalus*). Examples of useful bacteria include, but are not limited to, *Escherichia, Enterobacter, Azotobacter, Erwinia, Klebsielia.*

[0239] Typically, the polynucleotide that encodes the fusion protein is placed under the control of a promoter that is functional in the desired host cell. An extremely wide variety of promoters are well known, and can be used in the expression vectors of the invention, depending on the particular application. Ordinarily, the promoter selected depends upon the cell in which the promoter is to be active. Other expression control sequences such as ribosome binding sites, transcription termination sites and the like are also optionally included. Constructs that include one or more of these control sequences are termed "expression cassettes." Accordingly, the invention provides expression cassettes into which the nucleic acids that encode fusion proteins are incorporated for high level expression in a desired host cell.

[0240] Expression control sequences that are suitable for use in a particular host cell are often obtained by cloning a gene that is expressed in that cell. Commonly used prokaryotic control sequences, which are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences, include such commonly used promoters as the beta-lactamase (penicillinase) and lactose (*lac*) promoter systems (Change *et al.*, *Nature* (1977) 198: 1056), the tryptophan (*trp*) promoter system (Goeddel *et al.*, *Nucleic Acids Res.* (1980) 8: 4057), the *tac* promoter (DeBoer, *et al., Proc. Natl. Acad. Sci. U.S.A.* (1983) 80:21-25); and the lambda-derived P_L promoter and N-gene ribosome binding site (Shimatake *et al.*, *Nature* (1981) 292: 128). The particular promoter system is not critical to the invention, any available promoter that functions in prokaryotes can be used.

[0241] For expression of recombinant eukaryotic glycosyltransferases in prokaryotic cells other than *E. coli*, a promoter that functions in the particular prokaryotic species is required. Such promoters can be obtained from genes that have been cloned from the species, or heterologous promoters can be used. For example, the hybrid *trp-lac* promoter functions in *Bacillus* in addition to *E. coli*.

[0242] A ribosome binding site (RBS) is conveniently included in the expression cassettes of the invention. An RBS in *E. coli*, for example, consists of a nucleotide sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon (Shine and

Dalgarno, *Nature* (1975) 254: 34; Steitz, *In Biological regulation and development: Gene expression* (ed. R.F. Goldberger), vol. 1, p. 349, 1979, Plenum Publishing, NY).

[0243] For expression of the recombinant eukaryotic glycosyltransferases in yeast, convenient promoters include GAL1-10 (Johnson and Davies (1984) *Mol. Cell. Biol.* 4:1440-1448) ADH2 (Russell *et al.* (1983) *J. Biol. Chem.* 258:2674-2682), PHO5 (*EMBO J.* (1982) 6:675-680), and MFα (Herskowitz and Oshima (1982) in *The Molecular Biology of the Yeast Saccharomyces* (eds. Strathern, Jones, and Broach) Cold Spring Harbor Lab., Cold Spring Harbor, N.Y., pp. 181-209). Another suitable promoter for use in yeast is the ADH2/GAPDH hybrid promoter as described in Cousens *et al.*, *Gene* 61:265-275 (1987). For filamentous fungi such as, for example, strains of the fungi *Aspergillus* (McKnight *et al.*, U.S. Patent No. 4,935,349), examples of useful promoters include those derived from *Aspergillus nidulans* glycolytic genes, such as the ADH3 promoter (McKnight *et al.*, *EMBO J.* 4: 2093 2099 (1985)) and the *tpi*A promoter. An example of a suitable terminator is the ADH3 terminator (McKnight *et al.*).

[0244] Suitable constitutive promoters for use in plants include, for example, the cauliflower mosaic virus (CaMV) 35S transcription initiation region and region VI promoters, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, and other promoters active in plant cells that are known to those of skill in the art. Other suitable promoters include the full-length transcript promoter from Figwort mosaic virus, actin promoters, histone promoters, tubulin promoters, or the mannopine synthase promoter (MAS). Other constitutive plant promoters include various ubiquitin or polyubiquitin promoters derived from, *inter alia*, Arabidopsis (Sun and Callis, *Plant J.*, 11(5):1017-1027 (1997)), the mas, Mac or DoubleMac promoters (described in united States Patent No. 5,106,739 and by Comai *et al.*, *Plant Mol. Biol.* 15:373-381 (1990)) and other transcription initiation regions from various plant genes known to those of skill in the art. Useful promoters for plants also include those obtained from Ti- or Ri-plasmids, from plant cells, plant viruses or other hosts where the promoters are found to be functional in plants. Bacterial promoters that function in plants, and thus are suitable for use in the methods of the invention include the octopine synthetase promoter, the nopaline synthase promoter, and the manopine synthetase promoter. Suitable endogenous plant promoters include the ribulose-1,6-biphosphate (RUBP) carboxylase small subunit (ssu) promoter, the (α-conglycinin promoter, the phaseolin promoter, the ADH promoter, and heat-shock promoters.

[0245] Either constitutive or regulated promoters can be used in the present invention. Regulated promoters can be advantageous because the host cells can be grown to high densities before expression of the fusion proteins is induced. High level expression of heterologous proteins slows cell growth in some situations. An inducible promoter is a promoter that directs expression of a gene where the level of expression is alterable by environmental or developmental factors such as, for example, temperature, pH, anaerobic or aerobic conditions, light, transcription factors and chemicals. Such promoters are referred to herein as "inducible" promoters, which allow one to control the timing of expression of the glycosyltransferase or enzyme involved in nucleotide sugar synthesis. For *E. coli* and other bacterial host cells, inducible promoters are known to those of skill in the art. These include, for example, the *lac* promoter, the bacteriophage lambda P_L promoter, the hybrid *trp-lac* promoter (Amann *et al.* (1983) *Gene* 25: 167; de Boer *et al.* (1983) *Proc. Nat'l. Acad. Sci. USA* 80: 21), and the bacteriophage T7 promoter (Studier *et al.* (1986) *J. Mol. Biol.*; Tabor *et al.* (1985) *Proc. Nat'l. Acad. Sci. USA* 82: 1074-8). These promoters and their use are discussed in Sambrook *et al.*, *supra*. A particularly preferred inducible promoter for expression in prokaryotes is a dual promoter that includes a *tac* promoter component linked to a promoter component obtained from a gene or genes that encode enzymes involved in galactose metabolism (*e.g.*, a promoter from a UDPgalactose 4-epimerase gene (*gal*E)). The dual *tac-gal* promoter, which is described in PCT Patent Application Publ. No. WO98/20111, provides a level of expression that is greater than that provided by either promoter alone.

[0246] Inducible promoters for use in plants are known to those of skill in the art (*see, e.g.*, references cited in Kuhlemeier *et al* (1987) *Ann. Rev. Plant Physiol.* 38:221), and include those of the 1,5-ribulose bisphosphate carboxylase small subunit genes of *Arabidopsis thaliana* (the "ssu" promoter), which are light-inducible and active only in photosynthetic tissue.

[0247] Inducible promoters for other organisms are also well known to those of skill in the art. These include, for example, the arabinose promoter, the *lacZ* promoter, the metallothionein promoter, and the heat shock promoter, as well as many others.

[0248] A construct that includes a polynucleotide of interest operably linked to gene expression control signals that, when placed in an appropriate host cell, drive expression of the polynucleotide is termed an "expression cassette." Expression cassettes that encode the fusion proteins of the invention are often placed in expression vectors for introduction into

the host cell. The vectors typically include, in addition to an expression cassette, a nucleic acid sequence that enables the vector to replicate independently in one or more selected host cells. Generally, this sequence is one that enables the vector to replicate independently of the host chromosomal DNA, and includes origins of replication or autonomously replicating sequences. Such sequences are well known for a variety of bacteria. For instance, the origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria. Alternatively, the vector can replicate by becoming integrated into the host cell genomic complement and being replicated as the cell undergoes DNA replication. A preferred expression vector for expression of the enzymes is in bacterial cells is pTGK, which includes a dual *tac-gal* promoter and is described in PCT Patent Application Publ. N0. WO98/20111.

[0249] It may also be desirable to add regulatory sequences which allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Regulatory systems in prokaryotic systems include the lac, tac, and trp operator systems. In yeast, the ADH2 system or GAL1 system may be used. In filamentous fungi, the TAKA α-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter may be used as regulatory sequences.

[0250] The construction of polynucleotide constructs generally requires the use of vectors able to replicate in bacteria. A plethora of kits are commercially available for the purification of plasmids from bacteria (*see*, for example, EasyPrepJ, FlexiPrepJ, both from Pharmacia Biotech; StrataCleanJ, from Stratagene; and, QIAexpress Expression System, Qiagen). The isolated and purified plasmids can then be further manipulated to produce other plasmids, and used to transfect cells. Cloning in *Streptomyces* or *Bacillus* is also possible.

[0251] Selectable markers are often incorporated into the expression vectors used to express the polynucleotides of the invention. These genes can encode a gene product, such as a protein, necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics or other toxins, such as ampicillin, neomycin, kanamycin, chloramphenicol, or tetracycline. Alternatively, selectable markers may encode proteins that complement auxotrophic deficiencies or supply critical nutrients not available from complex

media, *e.g.*, the gene encoding D-alanine racemase for Bacilli. Often, the vector will have one selectable marker that is functional in, *e.g.*, *E. coli*, or other cells in which the vector is replicated prior to being introduced into the host cell. A number of selectable markers are known to those of skill in the art and are described for instance in Sambrook *et al.*, *supra*. A preferred selectable marker for use in bacterial cells is a kanamycin resistance marker (Vieira and Messing, *Gene* 19: 259 (1982)). Use of kanamycin selection is advantageous over, for example, ampicillin selection because ampicillin is quickly degraded by β-lactamase in culture medium, thus removing selective pressure and allowing the culture to become overgrown with cells that do not contain the vector.

[0252] Construction of suitable vectors containing one or more of the above listed components employs standard ligation techniques as described in the references cited above. Isolated plasmids or DNA fragments are cleaved, tailored, and re-ligated in the form desired to generate the plasmids required. To confirm correct sequences in plasmids constructed, the plasmids can be analyzed by standard techniques such as by restriction endonuclease digestion, and/or sequencing according to known methods. Molecular cloning techniques to achieve these ends are known in the art. A wide variety of cloning and *in vitro* amplification methods suitable for the construction of recombinant nucleic acids are well-known to persons of skill. Examples of these techniques and instructions sufficient to direct persons of skill through many cloning exercises are found in Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology*, Volume 152, Academic Press, Inc., San Diego, CA (Berger); and Current Protocols in Molecular Biology, F.M. Ausubel *et al.*, eds., *Current Protocols*, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1998 Supplement) (Ausubel).

[0253] A variety of common vectors suitable for use as starting materials for constructing the expression vectors of the invention are well known in the art. For cloning in bacteria, common vectors include pBR322 derived vectors such as pBLUESCRIPT™, and λ-phage derived vectors. In yeast, vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp series plasmids) and pGPD-2. Expression in mammalian cells can be achieved using a variety of commonly available plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses).

[0254] The methods for introducing the expression vectors into a chosen host cell are not particularly critical, and such methods are known to those of skill in the art. For example, the expression vectors can be introduced into prokaryotic cells, including *E. coli*, by calcium chloride transformation, and into eukaryotic cells by calcium phosphate treatment or electroporation. Other transformation methods are also suitable.

[0255] Translational coupling may be used to enhance expression. The strategy uses a short upstream open reading frame derived from a highly expressed gene native to the translational system, which is placed downstream of the promoter, and a ribosome binding site followed after a few amino acid codons by a termination codon. Just prior to the termination codon is a second ribosome binding site, and following the termination codon is a start codon for the initiation of translation. The system dissolves secondary structure in the RNA, allowing for the efficient initiation of translation. See Squires, et. al. (1988), *J. Biol. Chem.* 263: 16297-16302.

[0256] The recombinant eukaryotic glycosyltransferases of the invention can also be further linked to other bacterial proteins. This approach often results in high yields, because normal prokaryotic control sequences direct transcription and translation. In *E. coli*, *lacZ* fusions are often used to express heterologous proteins. Suitable vectors are readily available, such as the pUR, pEX, and pMR100 series (*see, e.g.*, Sambrook *et al., supra.*). For certain applications, it may be desirable to cleave the non-glycosyltransferase and/or accessory enzyme amino acids from the fusion protein after purification. This can be accomplished by any of several methods known in the art, including cleavage by cyanogen bromide, a protease, or by Factor X_a (*see, e.g.*, Sambrook *et al., supra.*; Itakura *et al., Science* (1977) 198: 1056; Goeddel *et al., Proc. Natl. Acad. Sci. USA* (1979) 76: 106; Nagai *et al., Nature* (1984) 309: 810; Sung *et al., Proc. Natl. Acad. Sci. USA* (1986) 83: 561). Cleavage sites can be engineered into the gene for the fusion protein at the desired point of cleavage.

[0257] More than one recombinant eukaryotic glycosyltransferase may be expressed in a single host cell by placing multiple transcriptional cassettes in a single expression vector, or by utilizing different selectable markers for each of the expression vectors which are employed in the cloning strategy.

[0258] A suitable system for obtaining recombinant proteins from *E. coli* which maintains the integrity of their N-termini has been described by Miller *et al. Biotechnology* 7:698-704 (1989). In this system, the gene of interest is produced as a C-terminal fusion to the first 76

residues of the yeast ubiquitin gene containing a peptidase cleavage site. Cleavage at the junction of the two moieties results in production of a protein having an intact authentic N-terminal reside.

[0259] The expression vectors of the invention can be transferred into the chosen host cell by well-known methods such as calcium chloride transformation for *E. coli* and calcium phosphate treatment or electroporation for mammalian cells. Cells transformed by the plasmids can be selected by resistance to antibiotics conferred by genes contained on the plasmids, such as the *amp, gpt, neo* and *hyg* genes.

VI. Proteins and protein purification

[0260] The recombinant eukaryotic glycosyltransferase proteins can be purified according to standard procedures of the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis and the like (*see*, generally, R. Scopes, *Protein Purification*, Springer-Verlag, N.Y. (1982), Deutscher, *Methods in Enzymology Vol. 182: Guide to Protein Purification.*, Academic Press, Inc. N.Y. (1990)). In preferred embodiments, purification of the recombinant eukaryotic glycosyltransferase proteins occurs after refolding of the protein. Substantially pure compositions of at least about 70 to 90%, homogeneity are preferred; more preferably at least 91%, 92%, 93%, 94%, 95%, 96%, or 97%; and 98 to 99% or more homogeneity are most preferred. The purified proteins may also be used, *e.g.*, as immunogens for antibody production.

[0261] To facilitate purification of the recombinant eukaryotic glycosyltransferase proteins of the invention, the nucleic acids that encode the recombinant eukaryotic glycosyltransferase proteins can also include a coding sequence for an epitope or "tag" for which an affinity binding reagent is available, *i.e.* a purification tag. Examples of suitable epitopes include the myc and V-5 reporter genes; expression vectors useful for recombinant production of fusion proteins having these epitopes are commercially available (*e.g.*, Invitrogen (Carlsbad CA) vectors pcDNA3.1/Myc-His and pcDNA3.1/V5-His are suitable for expression in mammalian cells). Additional expression vectors suitable for attaching a tag to the fusion proteins of the invention, and corresponding detection systems are known to those of skill in the art, and several are commercially available (e.g., FLAG" (Kodak, Rochester NY). Another example of a suitable tag is a polyhistidine sequence, which is capable of binding to metal chelate affinity ligands. Typically, six adjacent histidines are used, although one can use more or less than six. Suitable metal chelate affinity ligands that can serve as the binding

moiety for a polyhistidine tag include nitrilo-tri-acetic acid (NTA) (Hochuli, E. (1990) "Purification of recombinant proteins with metal chelating adsorbents" In Genetic Engineering: Principles and Methods, J.K. Setlow, Ed., Plenum Press, NY; commercially available from Qiagen (Santa Clarita, CA)).

[0262] Purification tags also include maltose binding domains and starch binding domains. Purification of maltose binding domain proteins is known to those of skill in the art. Starch binding domains are described in WO 99/15636, herein incorporated by reference. Affinity purification of a fusion protein comprising a starch binding domain using a betacylodextrin (BCD)-derivatized resin is described in USSN 60/468,374, filed May 5, 2003, herein incorporated by reference in its entirety.

[0263] Other haptens that are suitable for use as tags are known to those of skill in the art and are described, for example, in the Handbook of Fluorescent Probes and Research Chemicals (6th Ed., Molecular Probes, Inc., Eugene OR). For example, dinitrophenol (DNP), digoxigenin, barbiturates (see, e.g., US Patent No. 5,414,085), and several types of fluorophores are useful as haptens, as are derivatives of these compounds. Kits are commercially available for linking haptens and other moieties to proteins and other molecules. For example, where the hapten includes a thiol, a heterobifunctional linker such as SMCC can be used to attach the tag to lysine residues present on the capture reagent.

[0264] One of skill would recognize that modifications can be made to the glycosyltransferase catalytic or functional domains and/or accessory enzyme catalytic domains without diminishing their biological activity. Some modifications may be made to facilitate the cloning, expression, or incorporation of the catalytic domain into a fusion protein. Such modifications are well known to those of skill in the art and include, for example, the addition of codons at either terminus of the polynucleotide that encodes the catalytic domain to provide, for example, a methionine added at the amino terminus to provide an initiation site, or additional amino acids (*e.g.*, poly His) placed on either terminus to create conveniently located restriction enzyme sites or termination codons or purification sequences.

VII. Uses of refolded glycosyltransferases

[0265] The invention provides recombinant eukaryotic glycosyltransferase proteins and methods of using the recombinant eukaryotic glycosyltransferase proteins to enzymatically synthesize glycoproteins, glycolipids, and oligosaccharide moieties, and to glycoPEGylate

glycoproteins. The glycosyltransferase reactions of the invention take place in a reaction medium comprising at least one glycosyltransferase, acceptor substrate, and donor substrate, and typically a soluble divalent metal cation. In some embodiments, accessory enzymes and substrates for the accessory enzyme catalytic moiety are also present, so that the accessory enzymes can synthesize the donor substrate for the glycosyltransferase. The recombinant eukaryotic glycosyltransferase proteins and methods of the present invention rely on the use the recombinant eukaryotic glycosyltransferase proteins to catalyze the addition of a saccharide to an acceptor substrate.

[0266] A number of methods of using glycosyltransferases to synthesize glycoproteins and glycolipids having desired oligosaccharide moieties are known. Exemplary methods are described, for instance, WO 96/32491, Ito *et al.* (1993) *Pure Appl. Chem.* 65: 753, and US Patents 5, 352,670, 5,374,541, and 5,545,553.

[0267] The recombinant eukaryotic glycosyltransferase proteins prepared as described herein can be used in combination with additional glycosyltransferases, that may or may not have required refolding for activity. For example, one can use a combination of refolded recombinant eukaryotic glycosyltransferase protein and a bacterial glycosyltranferase, which may or may not have been refolded after isolation from a host cell. Similarly, the recombinant eukaryotic glycosyltransferase can be used with recombinant accessory enzymes, which may or may not be part of the fusion protein.

[0268] The products produced by the above processes can be used without purification. In some embodiments, oligosaccharides are produced. Standard, well known techniques, for example, thin or thick layer chromatography, ion exchange chromatography, or membrane filtration can be used for recovery of glycosylated saccharides. Also, for example, membrane filtration, utilizing a nanofiltration or reverse osmotic membrane as described in commonly assigned AU Patent No. 735695 may be used. As a further example, membrane filtration wherein the membranes have a molecular weight cutoff of about 1000 to about 10,000 can be used to remove proteins. As another example, nanofiltration or reverse osmosis can then be used to remove salts. Nanofilter membranes are a class of reverse osmosis membranes which pass monovalent salts but retain polyvalent salts and uncharged solutes larger than about 200 to about 1000 Daltons, depending upon the membrane used. Thus, for example, the oligosaccharides produced by the compositions and methods of the present invention can be retained in the membrane and contaminating salts will pass through.

VIII. Donor substrate/Acceptor substrates

[0269] Suitable donor substrates used by the recombinant glycosyltransferase fusion proteins and methods of the invention include, but are not limited to, UDP-Glc, UDP-GlcNAc, UDP-Gal, UDP-GalNAc, GDP-Man, GDP-Fuc, UDP-GlcUA, and CMP-sialic acid. Guo *et al.*, *Applied Biochem. and Biotech.* **68**: 1-20 (1997)

[0270] Suitable acceptor substrates used by the recombinant glycosyltransferase fusion proteins and methods of the invention include, but are not limited to, polysaccharides, oligosaccharides, proteins, lipids, gangliosides and other biological structures (*e.g.*, whole cells) that can be modified by the methods of the invention. Exemplary structures, which can be modified by the methods of the invention include any a of a number glycolipids, glycoproteins and carbohydrate structures on cells known to those skilled in the art as set forth is Table 1.

Table 1

Hormones and Growth Factors	Receptors and Chimeric Receptors
• G-CSF • GM-CSF • TPO • EPO • EPO variants • α-TNF • Leptin Enzymes and Inhibitors • t-PA • t-PA variants • Urokinase • Factors VII, VIII, IX, X • DNase • Glucocerebrosidase • Hirudin • α1 antitrypsin • Antithrombin III Cytokines and Chimeric Cytokines • Interleukin-1 (IL-1), 1B, 2, 3, 4 • Interferon-α (IFN-α) • IFN-α-2b • IFN-β • IFN-γ • Chimeric diptheria toxin-IL-2	• CD4 • Tumor Necrosis Factor (TNF) receptor • Alpha-CD20 • MAb-CD20 • MAb-alpha-CD3 • MAb-TNF receptor • MAb-CD4 • PSGL-1 • MAb-PSGL-1 • Complement • GlyCAM or its chimera • N-CAM or its chimera • LFA-3 • CTLA-IV Monoclonal Antibodies (Immunoglobulins) • MAb-anti-RSV • MAb-anti-IL-2 receptor • MAb-anti-CEA • MAb-anti-platelet IIb/IIIa receptor • MAb-anti-EGF • MAb-anti-Her-2 receptor Cells • Red blood cells • White blood cells (*e.g.*, T cells, B cells, dendritic cells, macrophages, NK cells, neutrophils, monocytes and the like • Stem cells

[0271] Examples of suitable acceptor substrates used in fucosyltransferase-catalyzed reactions, and examples of suitable acceptor substrates used in sialyltransferase-catalyzed reactions are described in Guo *et al.*, *Applied Biochem. and Biotech.* **68**: 1-20 (1997), but are not limited thereto.

IX. Glycosyltransferase reactions

[0272] The recombinant eukaryotic glycosyltransferase proteins, acceptor substrates, donor substrates and other reaction mixture ingredients are combined by admixture in an aqueous reaction medium. The medium generally has a pH value of about 5 to about 8.5. The selection of a medium is based on the ability of the medium to maintain pH value at the desired level. Thus, in some embodiments, the medium is buffered to a pH value of about 7.5. If a buffer is not used, the pH of the medium should be maintained at about 5 to 8.5, depending upon the particular glycosyltransferase used. For fucosyltransferases, the pH range is preferably maintained from about 6.0 to 8.0. For sialyltransferases, the range is preferably from about 5.5 to about 7.5.

[0273] Enzyme amounts or concentrations are expressed in activity units, which is a measure of the initial rate of catalysis. One activity unit catalyzes the formation of 1 μmol of product per minute at a given temperature (typically 37°C) and pH value (typically 7.5). Thus, 10 units of an enzyme is a catalytic amount of that enzyme where 10 μmol of substrate are converted to 10 μmol of product in one minute at a temperature of 37 °C and a pH value of 7.5.

[0274] The reaction mixture may include divalent metal cations (Mg^{2+}, Mn^{2+}). The reaction medium may also comprise solubilizing detergents (*e.g.*, Triton or SDS) and organic solvents such as methanol or ethanol, if necessary. The enzymes can be utilized free in solution or can be bound to a support such as a polymer. The reaction mixture is thus substantially homogeneous at the beginning, although some precipitate can form during the reaction.

[0275] The temperature at which an above process is carried out can range from just above freezing to the temperature at which the most sensitive enzyme denatures. That temperature range is preferably about 0°C to about 45°C, and more preferably at about 20°C to about 37°C.

[0276] The reaction mixture so formed is maintained for a period of time sufficient to obtain the desired high yield of desired oligosaccharide determinants present on oligosaccharide groups attached to the glycoprotein to be glycosylated. For large-scale preparations, the reaction will often be allowed to proceed for between about 0.5-240 hours, and more typically between about 1-18 hours.

[0277] One or more of the glycosyltransferase reactions can be carried out as part of a glycosyltransferase cycle. Preferred conditions and descriptions of glycosyltransferase cycles have been described. A number of glycosyltransferase cycles (for example, sialyltransferase cycles, galactosyltransferase cycles, and fucosyltransferase cycles) are described in U.S. Patent No. 5,374,541 and WO 9425615 A. Other glycosyltransferase cycles are described in Ichikawa *et al. J. Am. Chem. Soc.* 114:9283 (1992), Wong *et al. J. Org. Chem.* **57:** 4343 (1992), DeLuca, *et al., J. Am. Chem. Soc.* 117:5869-5870 (1995), and Ichikawa *et al.* In *Carbohydrates and Carbohydrate Polymers*. Yaltami, ed. (ATL Press, 1993).

[0278] Other glycosyltransferases can be substituted into similar transferase cycles as have been described in detail for the fucosyltransferases and sialyltransferases. In particular, the glycosyltransferase can also be, for instance, glucosyltransferases, *e.g.*, Alg8 (Stagljov *et al., Proc. Natl. Acad. Sci. USA* 91:5977 (1994)) or Alg5 (Heesen *et al. Eur. J. Biochem.* 224:71 (1994)), N-acetylgalactosaminyltransferases such as, for example, α(1,3) N-acetylgalactosaminyltransferase, β(1,4) N-acetylgalactosaminyltransferases (Nagata *et al. J. Biol. Chem.* 267:12082-12089 (1992) and Smith *et al. J. Biol Chem.* 269:15162 (1994)) and polypeptide N-acetylgalactosaminyltransferase (Homa *et al. J. Biol Chem.* 268:12609 (1993)). Suitable N-acetylglucosaminyltransferases include GnTI (2.4.1.101, Hull *et al., BBRC* 176:608 (1991)), GnTII, and GnTIII (Ihara *et al. J. Biochem.* 113:692 (1993)), GnTV (Shoreiban *et al. J. Biol. Chem.* 268: 15381 (1993)), O-linked N-acetylglucosaminyltransferase (Bierhuizen *et al. Proc. Natl. Acad. Sci. USA* 89:9326 (1992)), N-acetylglucosamine-1-phosphate transferase (Rajput *et al. Biochem J.* 285:985 (1992), and hyaluronan synthase. Suitable mannosyltransferases include α(1,2) mannosyltransferase, α(1,3) mannosyltransferase, β(1,4) mannosyltransferase, Dol-P-Man synthase, OCh1, and Pmt1.

[0279] For the above glycosyltransferase cycles, the concentrations or amounts of the various reactants used in the processes depend upon numerous factors including reaction conditions such as temperature and pH value, and the choice and amount of acceptor

saccharides to be glycosylated. Because the glycosylation process permits regeneration of activating nucleotides, activated donor sugars and scavenging of produced PPi in the presence of catalytic amounts of the enzymes, the process is limited by the concentrations or amounts of the stoichiometric substrates discussed before. The upper limit for the concentrations of reactants that can be used in accordance with the method of the present invention is determined by the solubility of such reactants.

[0280] Preferably, the concentrations of activating nucleotides, phosphate donor, the donor sugar and enzymes are selected such that glycosylation proceeds until the acceptor is consumed. The considerations discussed below, while in the context of a sialyltransferase, are generally applicable to other glycosyltransferase cycles.

[0281] Each of the enzymes is present in a catalytic amount. The catalytic amount of a particular enzyme varies according to the concentration of that enzyme's substrate as well as to reaction conditions such as temperature, time and pH value. Means for determining the catalytic amount for a given enzyme under preselected substrate concentrations and reaction conditions are well known to those of skill in the art.

X. Multienzyme oligosaccharide synthesis

[0282] As discussed above, in some embodiments, two or more enzymes may be used to form a desired oligosaccharide determinant on a glycoprotein or glycolipid. For example, a particular oligosaccharide determinant might require addition of a galactose, a sialic acid, and a fucose in order to exhibit a desired activity. Accordingly, the invention provides methods in which two or more enzymes, *e.g.*, glycosyltransferases, trans-sialidases, or sulfotransferases, are used to obtain high-yield synthesis of a desired oligosaccharide determinant.

[0283] In a particularly preferred embodiment, one of the enzymes used is a sulfotransferase which sulfonates the saccharide or the peptide. Even more preferred is the use of a sulfotransferase to prepare a ligand for a selectin (Kimura *et al., Proc Natl Acad Sci U S A* **96**(8):4530-5 (1999)).

[0284] In some cases, a glycoprotein- or glycolipid linked oligosaccharide will include an acceptor substrate for the particular glycosyltransferase of interest upon *in vivo* biosynthesis of the glycoprotein or glycolipid. Such glycoproteins or glycolipids can be glycosylated using the recombinant glycosyltransferase fusion proteins and methods of the invention without prior modification of the glycosylation pattern of the glycoprotein or glycolipid,

respectively. In other cases, however, a glycoprotein or glycolipid of interest will lack a suitable acceptor substrate. In such cases, the methods of the invention can be used to alter the glycosylation pattern of the glycoprotein or glycolipid so that the glycoprotein-or glycolipid-linked oligosaccharides then include an acceptor substrate for the glycosyltransferase-catalyzed attachment of a preselected saccharide unit of interest to form a desired oligosaccharide moiety.

[0285] Glycoprotein- or glycolipid linked oligosaccharides optionally can be first "trimmed," either in whole or in part, to expose either an acceptor substrate for the glycosyltransferase or a moiety to which one or more appropriate residues can be added to obtain a suitable acceptor substrate. Enzymes such as glycosyltransferases and endoglycosidases are useful for the attaching and trimming reactions. For example, a glycoprotein that displays "high mannose"-type oligosaccharides can be subjected to trimming by a mannosidase to obtain an acceptor substrate that, upon attachment of one or more preselected saccharide units, forms the desired oligosaccharide determinant.

[0286] The methods are also useful for synthesizing a desired oligosaccharide moiety on a protein or lipid that is unglycosylated in its native form. A suitable acceptor substrate for the corresponding glycosyltransferase can be attached to such proteins or lipids prior to glycosylation using the methods of the present invention. *See, e.g.,* US Patent No. 5,272,066 for methods of obtaining polypeptides having suitable acceptors for glycosylation.

[0287] Thus, in some embodiments, the invention provides methods for *in vitro* sialylation of saccharide groups present on a glycoconjugate that first involves modifying the glycoconjugate to create a suitable acceptor.

XI. Conjugation of modified sugars to peptides

[0288] The modified sugars are conjugated to a glycosylated or non-glycosylated peptide or protein using an appropriate enzyme to mediate the conjugation. Preferably, the concentrations of the modified donor sugar(s), enzyme(s) and acceptor peptide(s) or protein(s) are selected such that glycosylation proceeds until the acceptor is consumed. The considerations discussed below, while set forth in the context of a sialyltransferase, are generally applicable to other glycosyltransferase reactions.

[0289] A number of methods of using glycosyltransferases to synthesize desired oligosaccharide structures are known and are generally applicable to the instant invention.

Exemplary methods are described, for instance, WO 96/32491, Ito *et al.*, *Pure Appl. Chem.* 65: 753 (1993), and U.S. Pat. Nos. 5,352,670, 5,374,541, and 5,545,553.

[0290] In a some embodiments, an endoglycosidase is used in the reaction in combination with glycosyltransferases. The enzymes are used to alter a saccharide structure on the peptide at any point either before or after the addition of the modified sugar to the peptide.

[0291] In another embodiment, the method makes use of one or more exo- or endoglycosidase. The glycosidase is typically a mutant, which is engineered to form glycosyl bonds rather than rupture them. The mutant glycanase typically includes a substitution of an amino acid residue for an active site acidic amino acid residue. For example, when the endoglycanase is endo-H, the substituted active site residues will typically be Asp at position 130, Glu at position 132 or a combination thereof. The amino acids are generally replaced with serine, alanine, asparagine, or glutamine.

[0292] The mutant enzyme catalyzes the reaction, usually by a synthesis step that is analogous to the reverse reaction of the endoglycanase hydrolysis step. In these embodiments, the glycosyl donor molecule (*e.g.*, a desired oligo- or mono-saccharide structure) contains a leaving group and the reaction proceeds with the addition of the donor molecule to a GlcNAc residue on the protein. For example, the leaving group can be a halogen, such as fluoride. In other embodiments, the leaving group is a Asn, or a Asn-peptide moiety. In yet further embodiments, the GlcNAc residue on the glycosyl donor molecule is modified. For example, the GlcNAc residue may comprise a 1,2 oxazoline moiety.

[0293] In a preferred embodiment, each of the enzymes utilized to produce a conjugate of the invention are present in a catalytic amount. The catalytic amount of a particular enzyme varies according to the concentration of that enzyme's substrate as well as to reaction conditions such as temperature, time and pH value. Means for determining the catalytic amount for a given enzyme under preselected substrate concentrations and reaction conditions are well known to those of skill in the art.

[0294] The temperature at which an above process is carried out can range from just above freezing to the temperature at which the most sensitive enzyme denatures. Preferred temperature ranges are about 0 °C to about 55 °C, and more preferably about 20 ° C to about 30 °C. In another exemplary embodiment, one or more components of the present method are conducted at an elevated temperature using a thermophilic enzyme.

[0295] The reaction mixture is maintained for a period of time sufficient for the acceptor to be glycosylated, thereby forming the desired conjugate. Some of the conjugate can often be detected after a few hours, with recoverable amounts usually being obtained within 24 hours or less. Those of skill in the art understand that the rate of reaction is dependent on a number of variable factors (*e.g*, enzyme concentration, donor concentration, acceptor concentration, temperature, solvent volume), which are optimized for a selected system.

[0296] The present invention also provides for the industrial-scale production of modified peptides. As used herein, an industrial scale generally produces at least one gram of finished, purified conjugate.

[0297] In the discussion that follows, the invention is exemplified by the conjugation of modified sialic acid moieties to a glycosylated peptide. The exemplary modified sialic acid is labeled with PEG. The focus of the following discussion on the use of PEG-modified sialic acid and glycosylated peptides is for clarity of illustration and is not intended to imply that the invention is limited to the conjugation of these two partners. One of skill understands that the discussion is generally applicable to the additions of modified glycosyl moieties other than sialic acid. Moreover, the discussion is equally applicable to the modification of a glycosyl unit with agents other than PEG including other water-soluble polymers, therapeutic moieties, and biomolecules.

[0298] An enzymatic approach can be used for the selective introduction of PEGylated or PPGylated carbohydrates onto a peptide or glycopeptide. The method utilizes modified sugars containing PEG, PPG, or a masked reactive functional group, and is combined with the appropriate glycosyltransferase or glycosynthase. By selecting the glycosyltransferase that will make the desired carbohydrate linkage and utilizing the modified sugar as the donor substrate, the PEG or PPG can be introduced directly onto the peptide backbone, onto existing sugar residues of a glycopeptide or onto sugar residues that have been added to a peptide.

[0299] An acceptor for the sialyltransferase is present on the peptide to be modified by the methods of the present invention either as a naturally occurring structure or one placed there recombinantly, enzymatically or chemically. Suitable acceptors, include, for example, galactosyl acceptors such as Galβ1,4GlcNAc, Galβ1,4GalNAc, Galβ1,3GalNAc, lacto-N-tetraose, Galβ1,3GlcNAc, Galβ1,3Ara, Galβ1,6GlcNAc, Galβ1,4Glc (lactose), and other

acceptors known to those of skill in the art (*see, e.g.*, Paulson *et al., J. Biol. Chem.* **253**: 5617-5624 (1978)).

[0300] In one embodiment, an acceptor for the sialyltransferase is present on the glycopeptide to be modified upon *in vivo* synthesis of the glycopeptide. Such glycopeptides can be sialylated using the claimed methods without prior modification of the glycosylation pattern of the glycopeptide. Alternatively, the methods of the invention can be used to sialylate a peptide that does not include a suitable acceptor; one first modifies the peptide to include an acceptor by methods known to those of skill in the art. In an exemplary embodiment, a GalNAc residue is added by the action of a GalNAc transferase.

[0301] In an exemplary embodiment, the galactosyl acceptor is assembled by attaching a galactose residue to an appropriate acceptor linked to the peptide, *e.g.*, a GlcNAc. The method includes incubating the peptide to be modified with a reaction mixture that contains a suitable amount of a galactosyltransferase (*e.g.*, galβ1,3 or galβ1,4), and a suitable galactosyl donor (*e.g.*, UDP-galactose). The reaction is allowed to proceed substantially to completion or, alternatively, the reaction is terminated when a preselected amount of the galactose residue is added. Other methods of assembling a selected saccharide acceptor will be apparent to those of skill in the art.

[0302] In yet another embodiment, glycopeptide-linked oligosaccharides are first "trimmed," either in whole or in part, to expose either an acceptor for the sialyltransferase or a moiety to which one or more appropriate residues can be added to obtain a suitable acceptor. Enzymes such as glycosyltransferases and endoglycosidases (*see*, for example U.S. Patent No. 5,716,812) are useful for the attaching and trimming reactions.

[0303] Methods for conjugation of modified sugars to peptides or proteins are found *e.g.*, in USSN 60/328,523 filed October 10, 2001; USSN 60/387,292, filed June 7, 2002; USSN 60/391,777 filed June 25, 2002; USSN 60/404,249 filed August 16, 2002; and PCT/US02/32263; each of which are herein incorporated by reference for all purposes.

EXAMPLES

Example 1: Refolding Rat Liver ST3GalIII Expressed in Bacteria.

Refolding rat liver GST-ST3GalIII fusion protein

[0304] Rat liver *N*-acetyllactosaminide α-2,3-sialyltransferase (ST3GalIII) was cloned into pGEX-KT-Ext vector and expressed as GST-ST3-Gal III inclusion bodies in *E.coli* BL21 cells. Inclusion bodies were refolded using a GSH/GSSG redox system. The refolded

enzyme, GST-ST3-GalIII, was active and transferred sialic acid to an LNnT sugar substrate and to asialylated glycoproteins, for example, transferrin and Factor IX.

Cloning ST3GalIII into pGEX-XT-KT vector

[0305] Rat liver ST3-GalIII gene was cloned into *Bam*H1 and *Eco*R1 sites of the pGEX-KT-Ext vector after PCR Amplification using the following primers:

Sense	Sial 5'Tm	5'-TTTGGATCCAAGCTACACTTACTCCAATGG
Antisense:	Sial 3' Whole	5'-TTTGAATTCTCAGATACCACTGCTTAAGTC

Expression of GST-ST3GalIII in *E. coli* BL21cells

[0306] pGEX-ST3GalIII, an expression vector comprising the ST3GalIII GST fusion, was transformed into chemically competent *E. coli* BL21 cells. Single colonies were picked, inoculated into five ml LB media with 100 μg/ml carbenicillin, and grown overnight at 37°C with shaking. The next day, one ml of overnight culture was transferred into one liter of LB media with 100 μg/ml carbenicillin. Bacteria were grown until to an OD_{620} of 0.7, then 150 μM IPTG (final) was added to the medium. Bacteria were grown at 37°C for one to two hours more, then shifted to room temperature and grown overnight with shaking. Cells were harvested by centrifugation; bacterial pellets were resuspended in PBS buffer and lysed using a French Press. Soluble and insoluble fractions were separated by centrifugation for thirty minutes at 10,000 RPM in a Sorvall, SS 34 rotor at 4°C.

Purification of the inclusion bodies

[0307] Fifty ml of Novagen's Wash buffer (20 mM Tris.HCl, pH 7.5, 10 mM EDTA, 1 % Triton X-100) was added to the insoluble fraction, *i.e.*, the inclusion bodies (IB's). The insoluble fraction was vortexed to resuspend the pellet. The suspended IB's were centrifuged and washed at least twice by resuspending in Wash Buffer as above. Clean precipitates (IB's) were recovered and were stored at -20 °C until use.

Refolding inclusion bodies

[0308] The IB's were weighed (144 mg) and dissolved in Genotech IBS buffer (1.44 ml). The resuspended IB's were incubated at 4 °C for one hour in an Eppendorf centrifuge tube. Insoluble material was removed by centrifugation at maximum speed in an Eppendorf centrifuge. Solubilized IB's were diluted to 4 ml final volume. Refolding of GST-ST3GalIII was tested in refolding buffer solutions containing cyclodextrin, polyethylene glycol (PEG), ND SB-201, or a GSH/GSSG redox system. One ml of solubilized IB's were diluted rapidly by pipetting into the refolding solution, vigorously mixed for 30-40 seconds, and then gently

stirred for two hours at 4 °C. Three ml aliquots of the refolded GST-ST3GalIII solutions were dialyzed against cold PBS buffer or a buffer containing 50 mM Tris.HCL, pH 7.0; 100 mM NaCL; and 1 % glycerol using Pierce Slide-A-lyzers (MWCO:3.5 kDa,). After dialysis, the GST-ST3GalIII solutions were concentrated 3, 6 and 12 fold using Vivaspin 5 K (VivaScience) concentrators in Jouan centrifuge at 4,000 rpm at 4°C.

[0309] After refolding and dialysis, the refolded GST-ST3GalIII proteins were analyzed by SDS-polyacrylamide gel electrophoresis. The GST-ST3GalIII fusion, with a molecular weight of about 63-64 kDa, was present under all refolding conditions. (Data not shown.)

Sialylation of oligosaccharides using refolded GST-ST3 Gal III

[0310] Enzymatic assays using oligosaccharide substrates were carried out using CE-LIF (Capillary Electrophoresis-Laser Induced Fluorescence). Refolded ST3 Gal III enzymes were assayed for ability to transfer of sialic acid from CMP-NAN (cytidine 5-Monophosphate –β-D-sialic acid) to LNnT-APTS (Lacto-*N*-Neotetraose-9-aminopyrene 1-4, 6 trisulfonic acid) to form LSTd-APTS (Lactosialic-Tetrasaccharide- d-APTS). Reactions were performed in 96 well microtiter plates in 100 µl of a buffer containing 20 mM MOPS, pH 6.5; 0.8 mM CMP-NAN; 22.1 mM LNnT; 25 µM LNnT-APTS; 2.5 mM $MnCl_2$. Reactions were started by addition of 20 µl of refolded ST3 Gal III at 30 °C for thirty minutes. Reactions were quenched with a 1 to 25 dilution with water. The diluted reaction was analyzed by CE-LIF using an N-CHO coated capillary according to manufacturer's guide. Activities were calculated as the ratio of the normalized peak areas of LNnT-APTS to LSTd-APTS. Results comparing different refolding conditions are shown in Table 2. Two additional experiments using the GSH/GSSG system are shown in Table 3.

Table 2. GST-ST3-Gal III activities after screening different folding systems. The proteins were assayed directly without concentration.

Cyclodextrin	PEG	ND SB-201	GSH/GSSG
0	0	0	7.8 U/L*

*Activities reported here are Units per L refolded enzyme.

Table 3. GST-ST3GalIII activities after two separate folding experiments using GSH/GSSG system.

GSH/GSSG	Conc	Activity
Refolding Trial 1	12x	182 U/L*
Refolding Trial 2	40x	531 U/L*

*Activities reported here are Units per L refolded enzyme

Sialylation of glycoproteins using refolded GST-ST3 Gal III

[0311] Twenty μL of asialylated Transferrin (2μg/μL) or asialylated Factor IX (2μg/μL), was added to fifty μL of a buffer containing 50mM Tris, pH 8.0; and 150 mM NaCl, with 10 μL of 100 mM $MnCl_2$; 10 μL of 200mM CMP-NAN; and 0.05% sodium azide. The reaction mixture was incubated with 30 μL refolded GST-ST3GalIII at 30°C overnight or longer with shaking at 250 rpm. After the reactions were stopped, the sialylated proteins were separated on pH 7-3 IEF (Isoelectric focusing gel, Invitrogen) and stained with Comassie Blue according to manufacturer's guideline. Both Transferrin and Factor IX were sialylated by GST-ST3GalIII. (Data not shown.)

Refolding a rat liver ST3GalIII fused to an MBP tag.

[0312] Rat liver ST3GalIII was cloned into pMAL-c2x vector and expressed as a maltose binding protein (MBP) fusion, MBP-ST3GalIII, in inclusion bodies of *E.coli* TB1 cells. The refolded MBP-ST3GalIII was active and transferred sialic acid to LNnT, a sugar substrate, and to asialylated glycoproteins, for example asialo-transferrin.

Cloning ST3GalIII into pMAL-c2x vector

[0313] The rat liver ST3-GalIII nucleic acid was cloned into *Bam*H1 and *Xba*I sites of the pMAL-c2x vector after PCR Amplification using the following primers:

Sense ST3BAMH1 5'-TAATGGATTCAAGCTACACTTACTCCAATGG
Antisense: ST3XBA1 5'-GCGCTCTAGATCAGATACCACTGCTTAAGT

[0314] Nucleotides encoding amino acids 28-374, *e.g.*, the stem region and catalytic domain of ST3GalIII, were fused to the MBP amino acid tag.

[0315] Three other truncations of ST3GalIII were constructed and fused to MBP. The three ST3Gal III (Δ73, Δ85, Δ86) inserts were isolated by PCR using the following 5' primers (ST3 BamH1 Δ73) TGTATCGGATCCCTGGCCACCAAGTACGCTAACTT; (ST3 BamH1 Δ85) TGTATCGGATCCTGCAAACCCGGCTACGCTTCAGCCAT; and (ST3

BamH1 Δ86) TGTATCGGATCCAAACCCGGCTACGCTTCAGCCAT) respectively, in pairs with the common 3' primer (ST3-Xho1-GGTCTCCTCGAGTCAGATACCACTGCTTAA). Each PCR product was digested with BamHI and XhoI, subcloned into BamHI-XhoI digested pCWin2-MBP Kanr vector, transformed into TB1 cells, and screened for the correct construct.

[0316] PCR reactions were carried out under the following conditions. One cycle at 95°C for 1 minute. One μl vent polymerase was added. Ten of the following cycles were performed: 94°C for 1 minute; 65°C for 1 minute; and 72°C for 1 minute. After a final ten minutes at 72°C, the reaction was cooled to 4°C.

[0317] All of the ST3GalIII truncations had activity after refolding. The experiments described below were performed using the MBP Δ73ST3GalIII truncation.

Expression of MBP-ST3GalIII in *E. coli* TB1 cells

[0318] The pMAL-ST3GalIII plasmid was transformed into chemically competent *E. coli* TB1 cells. Three isolated colonies containing TB1/pMAL-ST3GalIII construct were picked from the LB agar plates. The colonies were grown in five ml of LB media supplemented with 60 μg/ml carbenicillin at 37°C with shaking until the liquid cultures reached an OD_{620} of 0.7. Two one ml aliquots were withdrawn from each culture and used to inoculate fresh media with or without 500 μM IPTG (final). The cultures were grown at 37°C for two hours. Bacterial cells were harvested by centrifugation. Total cell lysates were prepared heating the cell pellets in the presence of SDS and DTT. IPTG induced expression of MBP-ST3GalIII. (Data not shown.)

Expression of MBP-ST3GalIII and Purification of the inclusion bodies:

[0319] A one ml aliquot of TB1/pMAL-ST3GalIII overnight culture was inoculated into 0.5 liter of LB media with 50 μg/ml carbenicillin and grown to an OD_{620} of 0.7. Expression of MBP-ST3GalIII was induced by addition of 0.5 mM IPTG, followed by overnight incubation at room temperature. The next day bacterial cells were harvested by centrifugation. Cell pellets were resuspended in a buffer containing 75 mM TrisHCl, pH 7.4; 100 mM NaCl; and 1 % glycerol. Bacterial cells were lyzed using a French Press. Soluble and insoluble fractions were separated by centrifugation for thirty minutes, 4°C, 10,000 rpm, Sorvall, SS 34 rotor). Soluble and insoluble fractions were separated by centrifugation for thirty minutes at 10,000RPM in a Sorvall, SS 34 rotor at 4°C.

Purification of the inclusion bodies and refolding of MBP-ST3GalIII using GSH/GSSG

[0320] The MBP-ST3GalIII inclusion bodies were purified and suspended using the same methods and buffers used for the GST-ST3GalIII fusion proteins described above. The MBP-ST3GalIII were refolded using the GSH/GSSG system described above. The refolded MBP-ST3GalIII enzymes were dialyzed against cold 65 mM Tris.HCL pH 7.5, 100 mM NaCl, 1 % glycerol using Pierce SnakeSkin Dialysis bag (MWCO:7 kDa). The refolded and dialyzed MBP-ST3GalIII were concentrated from 3-14 fold using Vivaspin 5 K (VivaScience) concentrators in Jouan centrifuge at 4,000 rpm at 4°C. The refolded MBP-ST3GalIII proteins were analyzed by SDS-Polyacrylamide gel electrophoresis. An 81 kDa MBP-ST3GalIII was detected. (Data not shown.)

MBP-ST3 Gal III enzymatic activity assays

[0321] Refolded MBP-ST3 Gal III enzymes were assayed for ability to transfer sialic acid from CMP-NAN to LNnT-APTS to form LSTd-APTS, as described above. The refolded MBP-ST3 Gal III enzymes were active and transferred sialic acid to LNnT-APTS to form LSTd-APTS. (Data not shown.)

[0322] Refolded MBP-ST3 Gal III enzymes were assayed for ability to transfer sialic acid from CMP-NAN to glycoproteins. Transfer of sialic acid to asialo-Transferrin was assayed as described above, for GST-ST3-GalIII enzymes. The refolded MBP-ST3 Gal III enzymes were active and transferred sialic acid to asialo-Transferrin. (Data not shown.) Although refolded GST-ST3 Gal III and MBP-ST3 Gal III enzymes had similar activities for transfer of sialic acid to a soluble oligosaccharide acceptor molecule, refolded MBP-ST3 Gal III enzymes were more active in transfer of sialic acid to a a glycoprotein acceptor molecule.

Additional assays of conditions for refolding MBP-ST3GalIII

[0323] MBP-ST3GalIII was refolded using the conditions shown in Figure 1. The buffer, redox couple and detergent (if used) were mixed before addition of solubilized IB's to start the refolding reaction. IB's were diluted 1/20. MBP-ST3GalIII refolding was also successful using with different redox couples, for example Cystamine2 HCl/Cysteine at molar ratios of ¼, 4/1, 1/10, or 5/5. (Data not shown.)

ST3 Gal III enzymatic activity assays

[0324] Refolded MBP-ST3 Gal III enzymes were assayed for ability to transfer sialic acid from CMP-NAN to LNnT-APTS to form LSTd-APTS, as described above. Results are shown in Figure 1. The highest refolded MBP-ST3 Gal III activities were seen using

conditions, 8, 11, 13 and 16. When refolding was scaled up to five ml, MBP-ST3 Gal III proteins refolded using conditions 8 and 16 had the highest activity. (See, *e.g.*, Table 4.)

Table 4.

Condition	U/ L folded protein	U/g IB's
8	70	37.0
6	50	40.5

Purification of MBP-ST3GalIII on amylose column

[0325] Refolded MBP-ST3GalIII proteins from the 5 ml refolding preperation were combined and dialyzed against 100 mM TrisHCl pH 7.4, 100 mM NaCl and 1 % glycerol. The refolded MBP-ST3GalIII proteins were applied to an amylose column. Most of the refolded MBP-ST3GalIII protein was bound to the amylose column and eluted with 10 mM maltose. An elution profile is shown in Figure 2. Enzymatic activity of the MBP-ST3GalIII fractions was determined using the LnNT assay and is shown in Figure 3.

GlycoPEGYlation of asialotransferrin with refolded MBP-ST3GalIII:

[0326] Asialo-transferrin (2 mg/ml) was incubated with purified fractions of refolded 100 µl of MBP-ST3GalIII in the presence of CMP-SA-PEG (10 kDa, 1.6 mM) or CMP-SA-PEG (20 kDa, 1.06 mM) in 230 µl reaction. GlycoPEgylation reactions were carried out at 30°C overnight or for three days. Aliquots were withdrawn from the reactions and analyzed on 4-20 % SDS-polyacrylamide gel. Results are shown in Figure 4. Purified, refolded MBP-ST3GalIII transfers 10 or 20 K PEGylated sialic acids to asialo-transferrin.

Large scale MBP-ST3GalIII refolding

[0327] The following method was used to make large scale refolded MBP-ST3GalIII.

[0328] Wet IB's (470 mg) were dissolved IB solubilization Buffer (13 ml) in 15 ml culture tube. IB solubilization buffer includes the following: 4 M Guanidine HCl; 100 mM TrisHCl, pH 9; and 100 mM NaCl. IB's were incubated in IB solubilization buffer at 4°C for about 1 hour with gentle shaking. Any insoluble material was removed by centrifugation in 1.5 mL Eppendorf tubes, at 4°C at max speed, for 30 minutes. The solubilized IB's were transferred to clean tubes and protein concentration was determined using absorbance at 280 nm.

[0329] The following refolding solution was prepared and kept at 4°C: 55 mM MES buffer, pH 6.5; 264 mM NaCl; 11 mM KCl; 0.055 % PEG 550; 550 mM Arginine. The

buffer was supplemented with 0.3 mM Lauryl maltoside (LM); 0.1 mM oxidized glutathione (GSSG); 1 mM reduced glutathione (GSH) immediately before the addition of solubilized IB's. Two ml of solubilized IB's were added into 43 ml of refolding buffer in 50 ml sterile culture tube. The tube was placed on a rocker-shaker and gently shaken for 24 hours at 4°C. The refolded protein was dialyzed in dialysis tubing (MWCO: 7 kD) against Dialysis Buffer (100 mM Tris HCl, pH 7.5; 100 mM NaCl; and 5 % glycerol) twice (in 10-20 volume excess buffer).

[0330] The large scale dialyzed, refolded MBP-Gal III was analyzed for ST3GalIII activity, and exhibited about 53.6 U/g IB.

Example 2: Site Directed Mutagenesis of Human GnTI to Enhance Refolding.

[0331] A truncated human N-acetylglucosaminyltransferase I (103 amino terminal amino acids deleted) was expressed in *E.coli* as a maltose binding fusion protein (GnTI/MBP). The fusion protein was insoluble and was expressed in inclusion bodies. After solubilization and refolding, the GnTI/MBP fusion protein had low activity. The crystal structure of a truncated form of rabbit GnTI (105 amino terminal amino acids deleted) shows an unpaired cysteine residue (CYS123) near the active site. (See, *e.g.*, Unligil *et al.*, *EMBO J.* 19:5269-5280 (2000)). The corresponding unpaired cysteine in the human GnTI was identified as CYS121 and was replaced with a series of amino acids that are similar in size and chemical characteristics. The amino acids used include serine (Ser), threonine (Thr), alanine (Ala) and aspartic acid (Asp). In addition, a double mutant, ARG120ALA, CYS121HIS, was also made. The mutant GnTI/MBP fusion proteins were expressed in *E. coli*, refolded and assayed for GnTI activity towards glycoproteins.

[0332] Mutagenesis was done using a Quick Change Site-Directed Mutagenesis Kit from Stratagene. Additional restriction sites were introduced with some of the GnT1 mutations. For example an *Apa*I site (underlined, GGGCCCAC) was introduced into the GnT1 ARG120ALA, CYS121HIS mutant, *i.e.*, CGC CTG → **GCC CAC** (changes in bold). The following mutagenic oligonucleotides were used to make the double mutant: GnT1 R120A, C121H+, 5'CCGCAGCACTGTTCGGGCCCACCTGGACAAGCTGCTG 3'; and GnT1 R120A, C121H- 5'CAGCAGCTTGTCCAGGTGGGCCCGAACAGTGCTGCGG 3' (changes shown in bold). An *Asc*I site (underlined, GGCGCGCC) was introduced into the GnT1 CYS121ALA mutant, *i.e.*, CTG → **GCC** (changes in bold). The following mutagenic oligonucleotides were used to make the GnT1 CYS121ALA mutant: GnT1C123A+

5'AGCACTGTTCGGCGCGCCCTGGACAAGCTGCTG 3; and GnT1C123A-
5'CAGCAGCTTGTCCAGGGCGCGCCGAACAGTGCT 3'

[0333] The activity of the mutant proteins expressed in *E. coli* was compared to the activity of wild type GnT1 expressed in baculovirus. A CYS121SER GNTI mutant was active in a TLC based assay. In contrast, a CYS121THR mutant had no detectable activity and a CYS121ASP mutant had low activity. A CYS121ALA mutant was very active, and a double mutant, ARG120ALA, CYS121HIS, based on the amino acid sequence of the *C. elegans* GnT1 protein (Gly14), also exhibited activity, including transfer of GlcNAc to glycoproteins. Amino acid and encoding nucleic acid sequences of the GnT1 mutants are provided in Figures 7-11.

[0334] A second GnT1 truncation was made and fused to MBP: MBP-GnT1(D35). Figure 35 provides a schematic of the MBP-GnT1 fusion proteins, and depicts the truncations, *e.g.*, Δ103 or Δ35, and the Cys121Ser mutation (top). The bottom of the figure provides the full length human GnT1 protein. Mutations of Cys121 were also made in the MBP-GnT1(D35) protein.

[0335] Both fusion proteins were expressed in E. coli and both had activity for remodeling of the RNAse B glycoprotein. Figure 36 provides an SDS-PAGE gel showing in the right panel the refolded MBP-GnT1 fusion proteins: MBP-GnT1(D35) C121A, MBP-GnT1(D103) R120A + C121H, and MBP-GnT1(D103) C121A. The left panel shows the activities for remodeling the RNAse B glycoprotein of two different batches (A1 and A2) of refolded MBP-GnT1(D35) C121A at different time points. The MBP-GnT1 (D103) C121A also remodeled the RNAse B glycoprotein. Data not shown.

Example 3: MPB fusions to GalT1.

[0336] The following fusions between truncated bovine GalT1 and MBP were constructed: MBP-GalT1 (D129) wt, (D70) wt or (D129 C342T). (For the full length bovine sequence, see, *e.g.*, D'Agostaro *et al.*, *Eur. J. Biochem.* 183:211-217 (1989) and accession number CAA32695.) Each construct had activity after refolding. The amino acid sequence of the full length bovine GalT1 protein is provided in Figure 30. The mutants are depicted shematically in Figure 31 with a control protein GalT1(40) (S96A+C342T). *See, e.g.*, Ramakrishnan *et al.*, *J. Biol. Chem.* 276:37666-37671 (2001).

[0337] MBP-GalT1 (D70) was expressed in *E. coli* strain JM109. After overnight induction with IPTG, inclusion bodies were isolated from the insoluble pellet after cells were

lysed using a French Press. IB's were washed twice and then solubilized in 4 M GndHCl, 100 mM NaCl, 0.1 M TrisHCl pH 9.0. Refolding was done at pH 6.5 with GSSH/GSH (10/1) by dilution into refolding buffer (1/20, 0.1-0.2 mg/ml protein), followed by overnight incubation at 4°C, without shaking. Refolded proteins were dialyzed against 50 mM TrisHCl pH 8.0 twice (MWCO: 7 kD). The dialyzed MBP-GalT1 (D70) proteins were loaded onto an amylose column; washed; and then eluted with 10 mM Maltose.

[0338] GalT1 activity was assayed using oligosaccharides as an acceptor. The enzymatic assays were carried out using HPLC/PAD (High Performance Liquid Chromatography with Pulsed Amperometric Detection). The conversion of LNT2 (Lacto-N-Triose-2) into LNnT (Lacto-N-Neotetraose) using UDP-Gal (Uridine 5'-Diphosphogalactose) by GalTI enzyme was performed as follows: The reaction was carried out in 100 µl of 50 mM Hepes, pH 7 buffer containing 6 mM UDP-Gal, 5 mM LNT-2, 5 mM MnCl2 and 100 ul of refolded enzyme at 37°C for 60 minutes. The reaction was quenched (1 to 10 dilution) with water and centrifuged through a 10,000 MWCO spin filter. The filtrate was then diluted 1 to 10. This diluted reaction was analyzed by HPLC using a Dionex DX-500 system and a CarboPac PA1 column with sodium hydroxide buffer. The sample product peak area was compared to an LNnT calibration curve, and the activity was calculated based on the amount of LNnT produced per min per µl of enzyme in the reaction.

[0339] The purified MBP-GalT1 (D70) proteins had activity using both soluble oligosaccharides and glycoproteins (*e.g.*, RNAseB) as acceptor molecules. Results are shown in Figures 32 and 33. In the RNAse B remodeling assay, MBP-GalT1 (D70) was compared to a control protein GalT1(40) (S96A+C342T), which is an unfused truncation of the bovine GalT1 protein that was also expressed in *E. coli* and refolded. The MBP-GalT1 (D70) protein had more activity toward the RNAse B glycoprotein than did the GalT1(40) (S96A+C342T). The MBP-GalT1 (D129) truncation also had more activity toward the RNAse B glycoprotein than did the GalT1(40) (S96A+C342T) protein. (Data not shown.)

[0340] Kinetics of the refolded and purified MBP-GalT1 (D70) protein for glycosylation of RNAse B were determined and compared to NSO GalT1, a soluble form of the bovine GalT1 protein that was expressed in a mammalian cell system. As shown in Figure 34, the refolded and purified MBP-GalT1 (D70) had improved kinetics compared to the NSO GalT1 protein.

Example 4: One Pot Method of Refolding Multiple Glycosyltransferases.

[0341] Eukaryotic ST3GalIII, GalT1, and GnT1 enzymes build N-glycan chains on glycoproteins. Additional modifications, for example GlycoPEGylation, can be performed using CMP-NAN-PEG as a donor substrate. Eukaryotic ST3GalIII, GalT1, and GnT1 enzymes are typically expressed in eukaryotic expression systems, for example fungal or mammalian cells.

[0342] Eukaryotic ST3GalIII, GalT1, and GnT1 enzymes each fused to a maltose binding protein (MBP) domain were solubilized, combined, and refolded together in a single vessel. The MBP fused and refolded enzymes were active and were used to add N-glycans to glycoproteins or to glycoPEGylate glycoproteins. The refolding buffer included a redox couple, for example, glutathione oxidized/reduced (GSH/GSSG). Refolding was enhanced by addition of arginine and polyethylene glycol 3350 (PEG). The IB's can be solubilized individually and added to refolding buffer in different proportions or solubilized together from IB's and added to the refolding buffer directly. The one step purification or immobilization of these enzymes can also be done using the MBP fusion tag.

Preparation of a refolded glycosyltransferase mixture (SuperGlycoMix)

Preparation of the glycosyltransferases IB's

[0343] Bacterial strains used to produce eukaryotic ST3GalIII, GalT1, and GnT1 enzymes are shown in Table 5. The table also shows the estimated molecular weight of the MBP fusion proteins. (MW based on amino acid composition, Vector NTI software.) All nucleic acids encoding the eukaryotic enzymes were expressed from IPTG inducible expression vectors.

Table 5

Strain/Construct	Protein expressed (IBS's)	MW (kD)
JM109/pCWori-MBP-GaT1 (Δ129) C342T	MBP-GalT1(Δ129) C342T	74.2
JM109/pCWIN2-MBP-GnT1 (Δ103) C121A	MBP-GnT1(Δ103) C121A	82.4
TB1/pMAL-ST3GalIII	MBP-ST3GalIII	82

[0344] Following IPTG induction of *E. coli* cultures, IB's containing GnT1, GalT1 and ST3GalIII enzymes isolated by lysing the cells using a French Press or detergent lysis (Novagen's Bugbuster Reagent). Pellets were recovered after centrifugation and processed to obtain IB's, as described previously. IB's were washed at least two times using Novagen's

IB wash buffer. Washed IB's were stored at -20°C until they are ready to use in refolding experiments

[0345] IB's containing ST3GalIII, GalT1, or GnT1 were separately dissolved in a buffer containing 6 M Guanidine HCl, 50 mM TrisHCl pH 8.0, 5 mM EDTA, 10 mM DTT at 4°C for one hour. Cleared supernatants were obtained after centrifugation (Max speed at Eppendorf Micro-centrifuge). The protein content of the solubilized IB's was determined by measuring absorbance at 280 nm. The protein contents in Table 6 were determined based on the extinction coefficients of each MBP-Glycosyltransferase. The extinction coefficients were calculated using Vector NTi software (See Table 5)

Table 6. Protein concentrations in solubilized IB's.

Protein	A280 at 1 mg/ml	mg/ml
MBP-ST3GalIII	1.49	4.23
MBP-GalT1(Δ129) C342T	1.39	6.80
MBP-GnT1(Δ103) C121A	1.7	3.29

One pot refolding of Glycosyltyransferases

[0346] Solubilized IB's were mixed at equal amounts, as shown in Table 7.

Table 7. Solubilized IB's were mixed at following amounts before refolding.

Protein	V(mL)	mg	% of total protein
MBP-ST3GalIII	0.8	3.4	36
MBP-GalT1(Δ129) C342T	0.5	3.4	36
MBP-GnT1(Δ103) C121A	0.8	2.6	28
Total	2.1	9.4	100

[0347] The protein concentration of the total solubilized IB mixture was 4.5 mg/ml. The mixture was diluted approximately 1/20 in refolding buffer making the final concentration of the total protein mixture 0.22 mg/mL. Refolding buffer containing 55 mM MES, pH 6.5; 550 mM Arginine; 0.055 % PEG3350; 264 mM NaCl; 11 mM KCl; 1 mM GSH; and 0.1 mM GSSG. Refolding can also be performed in a buffer with Tris HCl, pH 8.2 and a Cysteine/Cystamine redox couple can be substituted for GSH/GSSG. The IB mixture was

diluted into the refolding buffer and incubated at 4°C overnight (16-18 hours). Estimated concentrations of the glycosyltransferases in refolding reaction:

MBP-ST3GalIII	0.081 mg/mL
MBP- GalT1 (Δ129) C342T	0.081 mg/mL
MBP-GnT1 (Δ103) C121A	0.062 mg/mL

[0348] After overnight refolding, the refolded glycosyltransferase mix was dialyzed to remove chaotropic agent (*i.e.* Guanidine HCl). Dialysis was carried out twice against 50 mM TrisHCl pH 8.0 at 4°C (20 fold per dialysis) in a dialysis bag (SnakeSkin, MWCO: 7 kD, Pierce). The dialyzed refolded glycosyltransferase mix (Superglycomix, SGM) was concentrated six fold using VivaSpin 6 mL (MWCO: 10 kD) centrifugal concentrators. After concentration, all three glycoproteins were present in the mixture, as determined by SDS-PAGE analysis. (Data not shown.) After concentrating the SGM, enzymatic activities of GnT1, GalT1, and ST3GalIII were determined.

Enzymatic activities of SuperGlycoMix

[0349] Superglycomix (SGM), the one pot refolded glycosyltransferase mix contains three glycosyltransferases: ST3GalIII, GalT1 and GnT1. These enzymes were individually assayed for their enzymatic activities and analyzed using the methods indicated below. The enzymatic activities are listed in Table 8.

ST3 Gal III enzymatic activity assays

[0350] ST3GalIII assays were carried out using HPLC/UV (High Performance Liquid Chromatography with Ultraviolet Detection). The conversion of LNnT (Lacto-*N*-Neotetraose) into LSTd (Lactosialic-Tetrasaccharide-d) using CMP-NAN (cytidine 5'-Monophosphate–β-D-sialic acid) by ST3GalIII enzyme was performed as follows. The reaction was carried out in a 96 well microtiter plate in 100 µl of 20 mM MOPS, pH 6.5 buffer containing 2 mM CMP-NAN, 30 mM LNnT, 10 mM $MnCl_2$ and 20 ul of refolded enzyme at 30°C for 120 minutes. The reaction was quenched by heating to 98°C for 1 min. The microtiter plate was centrifuged at 3600 rpm for 10 min to pellet any precipitate. 75 µl of supernatant was diluted 1:1 with 75 µl of water. The diluted reaction was analyzed by LC/UV using a YMC-Pack Polyamine II column with a sodium phosphate buffer/acetonitrile gradient and detection at 200 nm. The sample product peak area was compared to an LSTd calibration curve, and the activity was calculated based on the amount of LSTd produced per min per µl of enzyme in the reaction.

GalTI enzymatic activity assays:

[0351] The enzymatic assays were carried out using HPLC/PAD (High Performance Liquid Chromatography with Pulsed Amperometric Detection). The conversion of LNT2 (Lacto-*N*-Triose-2) into LNnT (Lacto-*N*-Neotetraose) using UDP-Gal (Uridine 5'-Diphosphogalactose) by GalTI enzyme was performed as follows. The reaction was carried out in 100 ul of 50 mM Hepes, pH 7 buffer containing 6 mM UDP-Gal, 5 mM LNT-2, 5 mM $MnCl_2$ and 100 µl of refolded enzyme at 37°C for 60 minutes. The reaction was quenched (1 to 10 dilution) with water and centrifuged through a 10,000 MWCO spin filter. The filtrate was then diluted 1 to 10. This diluted reaction was analyzed by HPLC using a Dionex DX-500 system and a CarboPac PA1 column with sodium hydroxide buffer. The sample product peak area was compared to an LNnT calibration curve, and the activity was calculated based on the amount of LNnT produced per min per µl of enzyme in the reaction.

GnTI enzymatic activity assays:

[0352] The activity of GnTI is determined by measuring the transfer of a tritiated sugar from UDP-^{3}H-GlcNAc (Uridine diphosphate N-acetyl-D-glucosamine [6-^{3}H(N)]) to n-octyl 3,6-Di-O-(α-mannopyranosyl) β-D-mannopyranoside (OM3), a trimannosyl core with an octyl tail. The reaction was carried out in 20 µl of 100 mM MES, pH 6.0 buffer containing 3 mM UDP-GlcNAc, 0.1 mM UDP-^{3}H-GlcNAc, 0.5 mM OM3, 20 mM $MnCl_2$ and 10 µl of refolded enzyme at 37°C for 60 minutes. The reaction was quenched (1 to 6 dilution) with water and applied to a polymeric reversed-phase resin in a 96 well format that was previously conditioned according to the manufacturer's recommendations. The resin was washed twice with 200 ul of water and the product was eluted with 50 µl of 100% MeOH into a capture plate. Scintillation fluid (200 µL) was added to each well and the plate was mixed and counted using a PerkinElmer TopCount NXT microplate scintillation counter. The activity was calculated based on the amount of ^{3}H-GlcNAc incorporated into the product per min per µl of enzyme in the reaction.

Table 8. Enzymatic activities of refolded Glycosyltransferases in SGM

Enzymatic activity	mU/mL
GnT1	1
GalT1	165
ST3GalIII	10

[0353] The activities reported in the table above are close or in the range when these enzymes were refolded separately. GnT1 and GalT1 activities are close to those obtained using mammalian or baculovirus expression systems. ST3GalIII activities are somewhat lower than in ST3GalIII preparation obtained after fungal expression system. The ST3GalIII assay used here is modified from the procedure and values reported here approximately 4-5 fold lower than those obtained a method based on CE-LIF (Capillary electrophoresis-Laser induced Fluorescence).

Remodeling RNAseB-Man$_5$ using Superglycomix

[0354] A small glycoprotein, RNAseB with one N linked Man5 sugar, was remodeled by SGM in the presence of UDP-sugars (UDP-GlcNAc and UDP-Gal). The remodeling reaction was carried out either using UDP-GlcNAc or both UDP-GlcNAc and UDP-Gal to test the both GnT1 and GalT1 activities. Eight µl of SGM was added to 10 mM MES buffer pH 6.5 containing 5 mM UDP-GlcNAc, or/and 5 mM UDP-Gal, 9 µg RNAseBMan$_5$, 5 mM $MnCl_2$ in 25 µl assay incubated at 33°C for overnight to 48 hours. At the end of the reaction, ten µl aliquots were dialyzed against H_2O and 1.5 µl samples were spotted on MALDI-TOF plates. Samples were analyzed on MALDI-TOF after being treated with TFA and cinnapinic acid.

[0355] The remodeling of RNAseBman5 was done by transferring GlcNAc and Gal on Man5 of the RNaseB. After 48 hrs incubation at 33°C, majority of the GlcNAc and Gal transfer onto RNAseB was accomplished as indicated in MALDI-TOF spectra of the remodeled RNAseBMan$_5$. Results are summarized in Table 9.

Table 9. MALDI-TOF Spectra of the species after SGM reactions.

	m/z		
	RNAseB		
Reaction	Man_5	Man_5-GlcNAc	Man_5GlcNAc-Gal
No Enzyme	14983	-	-
SGM+ UDP-GlcNAc	14973	15177	-
SGM+ UDP-GlcNAc +UDP-Gal	14982	15170	15348

GlycoPEGylation EPO remodeling using SGM

[0356] GlycoPEGylation (20 K) was carried out in one pot reaction composed of the following components: 10 mM MES pH 6.5, 5 mM $MgCL_2$, 5 mM UDP-GlcNAc, 5 mM UDP-GalNAc, 0.5 mM CMP- SA-PEG (20 kDa), 24 μg EPO, 8 μL concentrated SGM. In control reactions, SGM was replaced by individual enzymes either refolded or expressed in mammalian cells or insect cells or *Aspergillus*. After overnight incubations, the reactions were analyzed on SDS-polyacrylamide gel. Results are shown in Figure 5. SGM added 20K PEG to EPO.

Assessment of one pot refolding conditions for multiple glycosyltransferases

[0357] Conditions for refolding multiple glycosyltranferases were assessed, including pH and refolding two or three enzymes at once.

Preparation of glycosyltransferase inclusion bodies

[0358] *E. coli* strains transformed with glycosyltransferase expression plasmids were described previously, with one exception. MBP-ST3GalIII was expressed in JM109 cells from a pCWori-ST3GalIII plasmid. The inclusion bodies were isolated and solubilized as described above. Protein contents were assessed as described above and are shown in Table 10.

Table 10. Solubilized IB's were mixed at following amounts before refolding.

Protein	**A280**	**A280 (at 1 mg/ml)**	**mg**	**% (of sol. protein)**
MBP-ST3GalIII	32.3	1.49	21.7	13.6
MBP-GalT1(Δ129) C342T	35.7	1.39	25.7	13.7
MBP-GnT1(Δ103) C121S	42.8	1.7	25.2	9.7

One pot refolding of Glycosyltyransferase IB mixtures

[0359] After determining their protein contents, solubilized IB's were mixed at amounts shown before diluted in the refolding buffers (Table 11). Refolding experiments of the GT's were carried out in 44 ml volume at 4°C at stationary phase using buffer A or B (below) and 0.1 mM GSSG and 1 mM GSH. **Buffer A**: 55 mM MES pH 6.5, 550 mM Arginine, 0.055 % PEG3350, 264 mM NaCl, 11 mM KCl, supplemented with 1 mM GSH, 0.1 mM GSSG. **Buffer B**: 55 mM TrisHCl pH 8, 550 mM Arginine, 0.055 % PEG3350, 264 mM NaCl, 11 mM KCl, supplemented with 1 mM GSH, 0.1 mM GSSG.

Table 11. Mixing amounts of solubilized GT IB's in 2 mL IBSB

Refolding in Buffer A

Refold 1 (A-2x)	Conc(mg/mL)	V (mL)	mg
MBP-GnT1 (Δ103) C121S	25.2	0.2	5
MBP- GalT1 (Δ129) C342T	25.7	0.2	5
IBSB	-	1.6	-

Refold 2 (A-3x)	Conc(mg/mL)	V (mL)	mg
MBP-GnT1 (Δ103) C121S	25.2	0.2	5
MBP- GalT1 (Δ129) C342T	25.7	0.2	5
MBP-ST3GalIII	21.7	0.4	8.7
IBSB	-	1.2	-

Refolding in Buffer B

Refold 3 (B-2x)	Conc(mg/mL)	V (mL)	mg
MBP-GnT1 (Δ103) C121S	25.2	0.2	5
MBP- GalT1 (Δ129) C342T	25.7	0.2	5
IBSB	-	1.4	-

Refold 4 (B-3x)	Conc(mg/mL)	V (mL)	mg
MBP-GnT1 (Δ103) C121S	25.2	0.2	5
MBP- GalT1 (Δ129) C342T	25.7	0.2	5
MBP-ST3GalIII	21.7	0.4	8.7
IBSB	-	1.2	

[0360] For double refolding (2x, two glycosyltranferases) 10 mg total protein in 2 ml was added into 41 mL refolding buffer (above) 0.45 mL 100 mM GSH, 0.45 mL 10 mM GSSG, after dilution total protein was 0.44 mg/ml. For triple refolding (3x, three glycosyltransferases) 18.7 mg total protein in 2 ml was added into 41 mL refolding buffer (above), 0.45 mL 100 mM GSH, 0.45 mL 10 mM GSSG. After dilution total protein was 0.83

mg/ml. The protein concentrations were higher than previous triple refolding experiment (0.22 mg/ml in SGM). Estimated concentrations of the glycosyltransferases in refolding reaction follow:

MBP-ST3GalIII	0.39 mg/mL
MBP- GalT1 (Δ129) C342T	0.23 mg/mL
MBP-GnT1 (Δ103) C121S	0.23 mg/mL

[0361] After overnight refolding, the refolded glycosyltransferase mix was dialyzed. Dialysis was carried out twice against 50 mM TrisHCl pH 8.0 at 4°C in a dialysis bag (SnakeSkin, MWCO: 7 kD, Pierce). After dialysis, the glycosyltransferase mix was concentrated 9-12 fold using 6 mL VIVA-Spin (MWCO: 10 K) centrifugal concentrators.

[0362] SDS-PAGE analysis demonstrated that the proteins were present after refolding, dialysis, and concentration.

Enzymatic assays of refolded glycosyltransferase mixtures

[0363] Enzymatic assays were performed as described above. Results are shown in Table 12.

Table 12. Enzymatic activities of refolded Glycosyltransferases after double and triple refolding experiments.

Folding	Fold conc	Enzymatic activity	mU/mL
Buffer A (A-2x)		GnT1 GalT1	0.84 598
Buffer A (A-3x)		GnT1 GalT1 ST3GalIII	0.16 306 4
Buffer B (B-2x)		GnT1 GalT1	3.32 747
Buffer B (B-3x)		GnT1 GalT1 ST3GalIII	0.47 425 11

[0364] The highest activity was seen on mixing MBP fused GnT1 and GalT1 in equal amounts and refolded in buffer B. Adding non-equivalent amount of MBP-fused ST3GalIII affected refolding efficiency due to total high protein. Nevertheless, two different refolding buffer using either two GT's or three GT's, can be used to obtain active soluble proteins.

Example 5: Refolding eukaryotic GalNAcT2.

[0365] A truncated human GalNAcT2 enzyme was expressed in *E. coli* and used to determine optimal conditions for solubilization and refolding using the methods described above. The full length human GalNAcT2 nucleic acid and amino acid sequences are provided in Figures 13A and B. The sequences of the mutant protein, GalNAcT2(D51), are shown in Figures 14A and B. The mutant was expressed in *E. coli* as an MBP fusion protein, MBP-GalNAcT2(D51). Other GalNAcT2 mutants were made, expressed in E. coli and were able to be refolded: MBP-GalNAcT2(D40), MBP-GalNAcT2(D73), and MBP-GalNAcT2(D94). Data not shown. Details of the construction of the additional deletion mutants is found in USSN 60/576,530, filed June 3, 2004 and USSN 60/598,584, August 3, 2004, both of which are herein incorporated by reference for all purposes.

[0366] Cultures of bacteria expressing MBP-GalNAcT2(D51) were grown and harvested as described above. Inclusion bodies were purified from bacteria as described above. Solubilization of the inclusion bodies was performed at pH 6.5 or at pH 8.0. After solubilization, MBP-GalNAcT2(D51) protein was refolded at either pH 6.5 or pH 8.0 using buffers A and B, *i.e.*, **Buffer A**: 55 mM MES pH 6.5, 550 mM Arginine, 0.055 % PEG3350, 264 mM NaCl, 11 mM KCl, supplemented with 1 mM GSH, 0.1 mM GSSG; and **Buffer B**: 55 mM TrisHCl pH 8, 550 mM Arginine, 0.055 % PEG3350, 264 mM NaCl, 11 mM KCl, supplemented with 1 mM GSH, 0.1 mM GSSG. After refolding, MBP-GalNAcT2(D51) protein was dialyzed and then concentrated. Figure 15 provides a demonstration of the protein concentration of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0.

[0367] A radiolabeled [^{3}H]-UDP-GalNAc assay was performed to determine the activity of the *E.coli*-expressed refolded MBP-GalNAcT2(D51) by monitoring the addition of radiolabeled GalNAc to a peptide acceptor. The acceptor was a MuC-2 – like peptide having the sequence MVTPTPTPTC). The peptide was dissolved in 1M Tris-HCl pH=8.0. See, *e.g.*, USSN 60/576,530 filed June 3, 2004; and US provisional patent application Attorney Docket Number 040853-01-5149-P1, filed August 3, 2004; both of which are herein incorporated by reference for all purposes. Figure 16 provides a demonstration of the enzymatic activity of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. Figure 17 provides a demonstration of the specific activity of refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. The highest activity levels were observed with MBP-GalNAcT2(D51) that

had been solubilized at pH 8.0 and refolded at pH 8.0. The highest specific activity levels were also observed with solubilization at pH 8.0 and refolding at pH 8.0.

[0368] Solubilized and refolded MBP-GalNAcT2(D51) was assayed for its ability to add GalNAc to the G-CSF protein. The assay consisted of an aliquot of enzyme and a reaction buffer (27mM MES, pH=7, 200mM NaCl, 20mM MgCl2, 20mM MnCl2, and 0.1% Tween 80), G-CSF Protein (2mg/ml in H_2O), and 100mM UDP-GalNAc. For each refold sample, 4.4μL of sample were added to 15μL of reaction solution. For the positive control, 1μL of standard GalNAcT2 Baculovirus was added along with 3.4μL of H_2O to one tube. Reactions were incubated at 32°C on a rotary shaker for several days, during which time an overnight time point and a 5 day time point were assayed by MALDI. See, *e.g.*, USSN 60/576,530 filed June 3, 2004; and US provisional patent application Attorney Docket Number 040853-01-5149-P1, filed August 3, 2004; both of which are herein incorporated by reference for all purposes.

[0369] Figures 18A and 18B provide results of remodeling of recombinant granulocyte colony stimulating factor (GCSF) using refolded MBP-GalNAcT2(D51) after solubilization at pH 6.5 or pH 8.0 and refolding at pH 6.5 or pH 8.0. A positive control, *i.e.*, purified MBP-GalNAcT2(D51) that had been expressed in baculovirus, and a negative control, *i.e.*, reaction mixture lacking a substrate were included. The highest levels of GCSF remodeling activity were seen using MBP-GalNAcT2(D51) that had been solubilized at pH 8.0 and refolded at pH 8.0.

Example 6: Refolding and purification of eukaryotic GalNAcT2.

[0370] Four liters of bacteria that express recombinant MBP-GalNAcT2(D51) were grown and harvested. Inclusion bodies were isolated, washed, and two grams dry weight of inclusion bodies were solubilized at 4°C in 200mL of solubilization buffer (7M urea/ 50mM Tris/ 10mM DTT/ 5mM EDTA at pH 8.0). After solubilization, the mixture was then diluted in to 4L of refolding buffer (50mM Tris/ 550mM L-Arginine/ 250mM NaCl/ 10mM KCl/ 0.05% PEG 3350/ 4mM L-cysteine/ 1mM cystamine dihydrochloride at pH 8.0). Refolding was carried out at 4-10°C for about 20 hours, with stirring. The mixture was then filtered using a 10SP CUNO filter, concentrated 5 fold on 4ft2 membrane, diafiltered 4 times with 10mM Tris/ 5mM NaCl at pH 8.0. The conductivity of the final refolded MBP-GalNAcT2(D51) solution was 1.4 mS/cm. The refolded protein was stored at 4°C for several days.

[0371] The refolded proteins were applied to a Q Sepharose XL (QXL) column (Amersham Biosciences, Piscataway, NJ). An elution profile is shown in Figure 19 and the enzymatic activity of specific column fractions are shown in Figure 20. The active fractions were combined and applied to an Hydroxyapatite Type I (80μm) (BioRad, Hercules, CA) column. An elution profile is shown in Figure 21 and activity of HA type I eluted fractions is shown in Figure 22. The combination of QXL and HA type I chromatography resulted in active, highly purified MBP-GalNAcT2(D51).

Example 7: Purification of eukaryotic MBP-SBD ST3Gal3.

[0372] The Δ73 ST3GalIII truncation was fused in frame to the maltose binding protein and a starch binding domain to form a doubly tagged protein: MBP-SBD-ST3Gal3 (Δ73). The MBP was at the amino terminus, followed by the SBD, and then the truncated ST3GalIII protein. Refolding and purification of the MBP-SBD-ST3Gal3 (Δ73) protein was compared to a single tag protein: MBP-ST3GalIII (Δ73). Both proteins were expressed in *E. coli* as insoluble inclusion bodies, and were solubilized and refolded as described herein. The proteins were then dialyzed and subjected to affinity purification using a cyclodextrin column that binds to both the MBP and SBD tags. Results are shown in Figure 23. The MBP-SBD-ST3Gal3 (Δ73) protein had higher specific activity after dialysis and retained more specific activity after elution from the column with cyclodextrin.

Example 8: Refolding of MBP-ST3Gal1 and MBPSBD-ST3Gal1 proteins

[0373] Eukaryotic ST3Gal1 was fused to MPB or MBP and SBD. The DNA sequence of the porcine ST3Gal1 gene was used as a template for the design of the pcWINMBP-pST3Gal1 and pcWINMBP/SBD-pST3Gal1 constructs described herein. The full length porcine ST3Gal1 sequence is provided in Figure 37. For expression in *E.Coli*, a codon optimized and truncated version of pST3Gal1 was used, *i.e.*, pST3Gal1 Δ45. The encoded amino acid sequences are provided in Figure 24. The ST3Gal1 gene coding sequence was next digested and transferred to pcWIN2-MBP and pcWINMBP/SBD vectors using the BamHI and XhoI cloning sites. These constructs were confirmed to be correct by restriction and sequence analysis and then were used to transform the *E.Coli* strain JM109 using 50ug/ml Kanamycin selection. An individual colony from each was used to inoculate a 2ml culture of Maritone-10ug/ml Kanamycin that was incubated for 16hrs at 37°C. Each culture was mixed separately 1:1 with 50% glycerol and frozen at -80 °C and referred as the stock vial. A small amount of each stock vial was used to streak a Maritone-Kan plate. After 16hr

incubation at 37 °C, a single colony from each was used to inoculate a 25ml culture of Maritone-10ug/ml Kanamycin that was incubated for 16hrs at 37°C. The 25ml culture was then used to inoculate a 1 L culture of Maritone-10ug/ml Kanamycin that was incubated at 37°C and monitored for OD600. When the OD600 reached 0.8, IPTG was added to 1mM and the cells were incubated for an additional 16hrs. Cells were then harvested by centrifugation at 7000xG for 15mins.

[0374] Inclusion bodies were then isolated, the ST3Gal1 fusion proteins were solubilized, and refolded. Bacterial cells pellets were resuspended at a ratio of 1g wet cell pellet per 10 mL of 20mM Tris pH 8, 5mM EDTA and lysed by mechanical disruption with two passes through a microfluidizer. Insoluble material, *i.e.*, the inclusion bodies or IB's, was pelleted by centrifugation at 7000xg at 4°C in the Sorvall RC3 for 30 minutes, and the supernatant discarded. A typical wash cycle consisted of completely resuspending the pellet in the wash buffer, and repeating the centrifugation for 15 minutes. The pellet was washed once in an excess volume (at least 10 mL (up to 20) per g original cell pellet) of high salt buffer (wash I: 10 mM Tris pH 7.4, 1 M NaCl, 5 mM EDTA), once in detergent buffer (wash II: 25 mM Tris pH 8, 100 mM NaCl, 1% Triton X100, 1% Na-deoxycholate, 5 mM EDTA), and (to remove traces of detergent in addition to washing the IBs) three times in wash buffer (wash III: 10 mM Tris pH 8, 5 mM EDTA).

[0375] Washed IB's were resuspended in solubilization buffer (8M Urea, 50mM BisTris pH 6.5, 5mM EDTA, 10mM DTT) and adjusted to 2mg/ml protein concentration. Refolding was performed as a rapid 20 fold dilution into refold buffer (55mM Tris pH 8.2, 10.56 mM NaCl, 0.44mM KCl, 2.2 mM MgCl2, 2.2 mM CaCl2, 0.055% Peg3350, 550mM L-Arginine, 1mM GSH, 100mcM GSSG) and stirred for 16hrs at 4°C. The buffer was then exchanged by desalting using G50 Sephadex to 50mM BisTris pH 6.5, 75mM NaCl, 0.05% Tween 80.

[0376] To determine if the refolded enzymes were active, a sialyltransferase assay was performed. Transfer of sialic acid to the donor (asialo-bovine submaxillary mucin) was monitored using radiolabeled CMP-NAN. Chicken ST6GalNac1 expressed in a baculovirus system was used as a positive control.

[0377] Briefly, 40 µL of the reaction mix was added to 10 µL of enzyme sample and incubated at 37°C for 1hour. The glycoprotein was precipitated by adding 100 µL of phosphotungstic acid/15% TCA to the reaction with mixing. After centrifugation, supernatants were aspirated and discarded. Five hundred µL of 5% TCA was added to wash

unincorporated CMP-^{14}C-sialic acid from the pellet. Reactions were centrifuged again and supernatants were aspirated and discarded. Pellets were resuspended using 100 µL of 10 N NaOH. One mL of 1 M Tris Buffer, pH 7.5 was added to the resuspended pellet and then the mixture was transferred to a scintillation vial. Five mL of scintillation fluid (Ecolume, ICN Biomedicals) was added and mixed well. The total counts added to a reaction was determined by adding 40 µL of reaction mix into a scintillation vial and adding 100 µL of 10 N NaOH, 1 mL of water, and 5 mL of scintillation fluid, and mixing well. Vials were counted for 1 minute. The reaction conditions are provided in Table 13.

TABLE 13

Reagent	Manufacturer	Cat. No.	Amount per Assay Tube	Final Concentration in the Assay
CMP [^{14}C] Sialic Acid	Amersham	CFB-165	4 uL	100,000 CPM
7 mM CMP-Sialic Acid	Neose	AES 533pg.42	1.4 uL	0.2 mM
25 mg/mL Asialo-Bovine Submaxillary Mucin	Sigma, Hydrolyzed at Neose	M-3895	20 uL	250-500 ug
1 M Bis/Tris Buffer pH 6.5	Sigma	B-9754	2.5 uL	50 mM
5 M NaCl	Sigma	S-7653	1 uL	100 mM
dH2O	N/A	N/A	11.1 uL	

Total Reaction Mix Volume per Assay Tube: 40 uL

[0378] Results are shown in Figure 25. Both refolded fusion proteins, MBP-pST3Gal1 and MBP-SBD-pST3Gal1 had detectable sialyltransferase activity.

Example 9: Refolding of MBP-ST6GalNAc1 proteins

[0379] Eukaryotic ST6GalNAc I was fused to MPB. Briefly, five mouse ST6GalNAc I constructs were generated: D32, E52, S127, S186, and S201. Each construct was expressed behind the MBP-tag from the vector pcWin2-MBP, and differ in the extent of the 'stem' region included in the construct. D32 is the longest form, starting immediately downstream of the predicted amino-terminal transmembrane domain. S201 is the shortest, beginning shortly before the predicted start of the conserved catalytic domain.

[0380] In addition to the mouse constructs, human ST6GalNAc I K36 was also expressed as a fusion with MBP. The human construct begins just after the transmembrane domain.

DNA encoding human ST6GalNAc1 from K36 to its c-terminus was isolated by PCR using the existing baculovirus expression vector as template, and cloned into the BamHI-XhoI sites within pcWin2-MBP.

[0381] For reference, the sequences for MBP-mST6GalNAcI S127 and MBP-hST6GalNAcI K36are included in Figure 26. In addition, Figure 38 provides full length amino acid sequences for human ST6GalNAcI and for chicken ST6GalNAcI, and a sequence of the mouse ST6GalNAcI protein beginning at residue 32 of the native mouse protein.

[0382] Deletion mutants additional to those described above have been made and a complete list of preferred ST6GalNAcI for use in the invention is found is Table 14. Figure 39 provides a schematic of a number of preferred human ST6GalNAcI truncation mutants. Figure 40 shows a schematic of MBP fusion proteins including the human ST6GalNAcI truncation mutants.

Table 14: ST6GalNAcI Mutants

	Truncation Site	Mutation
HUMAN	Δ35	K36
	Δ124	K125
	Δ257	S258
	Δ35	K36
	Δ72	T73
	Δ109	E110
	Δ133	M134
	Δ170	T171
	Δ232	A233
	Δ272	G273
CHICKEN		
	Δ48	Q49
	Δ152	V153
	Δ225	L226
	Δ232	T233
MOUSE	Δ31	D32
	Δ51	E52
	Δ126	S127
	Δ185	S186
	Δ200	S201

[0383] Figure 45 shows the position of paired and unpaired cysteine residues in the human ST6GalNAcI protein. Single and double cysteine substitution are also shown, *e.g.*, C280S, C362S, C362T, (C280S + C362S), and (C280S + C362T).

[0384] Initial expression studies showed that the ST6GalNAcI fusion proteins were expressed as insoluble proteins. To recover active recombinant enzyme, the inactive, insoluble proteins were isolated and refolded as described:

[0385] Logarithmically growing 0.5L cultures of JM109 cells bearing either pcWin2-MBP-mST6GalNAcI D32, E52, S127, S186, or pcWin2-MBP-hST6GalNAcI K36 were induced with 1 mM IPTG overnight at 37°C. Cells were collected by centrifugation, and lysed by mechanical disruption in a microfluidizer in 100 mL of 20 mM Tris pH8, 5 mM EDTA. Insoluble matter was collected by centrifugation at 7000 x g for 20 minutes. The supernatants were discarded, and the pellets were washed with a high salt buffer (20 mM Tris pH 7.4, 1M NaCl, 5 mM EDTA), detergent buffer (25 mM Tris pH 8, 1% Na-deoxycholate, 1% Triton x100, 100 mM NaCl, 5 mM EDTA), and TE (10 mM Tris pH 8, 1mM EDTA). Each wash was in 100 mL, and the pellet was collected by centrifugation as described above. Following the washing, the inclusion body pellets were aliquoted and stored at -80°C.

[0386] To screen for conditions that allow proper refolding and thus recovery of ST6GalNAc I activity, aliquots of the mouse and human ST6GalNAcI fusion protein inclusion bodies were solubilized in 6M guanidine, 10 mM DTT, 1x TBS. Protein concentration was normalized by Bradford assay, and the solubilized proteins were transferred to a series of commercially-available protein refolding buffers. Refolds were carried out in 0.25 mL at 0.2 mg/mL overnight at 4°C in a 96-well plate with shaking. The refolds were transferred to a 96-well dialysis plate (25000 MWCO) and dialyzed against 1x TBS, 0.05% Tween-80 for four hours at 4°C, followed by overnight dialysis against 10 mM BisTris pH 7.1, 100 mM NaCl, 0.05% Tween-80 at 4°C.

[0387] Refolded recombinant ST6GalNAcI fusion proteins were tested for activity in a 384-well solid phase activity assay. Briefly, the activity assay detects the ST6GalNAcI-mediated transfer of a biotinylated sialic acid from biotinylated CMP-NAN to the surface of an asialo-bovine submaxillary mucin-coated well in a 384-well plate. Each reaction (13.5 µL refold + 1.5 µL 10x reaction buffer) was performed in quadruplicate. 10x reaction buffer was 0.2M BisTris ph 6.7, 25 mM MgCl2, 25 mM MnCl2, 0.5% Tween-80, and 1 mM donor. After overnight incubation at 37°C, the plate was washed with excess 1x TBS, 0.05% Tween-

20, and biotin detected with europium-labeled streptavidin as per manufacturer's instructions (Perkin Elmer). Europium fluorescence levels retained on the plate, indicative of ST6GalNAcI activity, were documented with a Perkin Elmer Victor3V plate reader, and expression and activity results are summarized in Table 15. Three of the refolded ST6GalNAcI fusion proteins had detectable activity.

Table 15: Summary of refolded ST6GalNAcI fusion proteins tested for activity by solid phase assay.

Construct	Refolded protein detected by SDS-PAGE	Refolded protein activity detected by solid phase assay
MBP-mST6GalNAcI D32	+	-
MBP-mST6GalNAcI E52	++	-
MBP-mST6GalNAcI S127	+++	+
MBP-mST6GalNAcI S186	+/-	+/-
MBP-hST6GalNAcI K36	+/-	+

Example 10: Refolding of Core 1 GalT1 proteins

[0388] Eukaryotic Core 1 GalT1 is fused to MPB or to the double tag, MBPSPD. The drosophila and human Core 1 GalT1 proteins are used. Figure 41 provides the full length sequence of human Core 1 GalT1 protein. Figure 42 provides the sequences of two drosophila Core 1 GalT1 proteins. Truncations of each enzyme are made throughout the stem region, *i.e.*, starting with the full length stem region and deleting one amino acid at a time such that the smallest truncation comprises only a Core 1 GalT1 catalytic domain. Cysteine residues throughout the proteins' catalytic domain are also mutated one at a time to either serine or alanine residues. MBP fusions are made using the truncated proteins, the cysteine mutants or combinations of the two. Proteins are expressed in *E. coli* as inclusion bodies, are solubilized, and are then refolded using the methods described herein. Refolding is determined by measuring enzymatic activity. Active enzymes have been correctly refolded. Enzymatic activity of Core 1 GalT1 is measured as disclosed in Ju *et al.*, *J. Biol. Chem.* 277:178-186 (2002), which is herein incorporated by reference for all purposes.

Example 11: One pot refolding of O-linked glycosyltransferases

[0389] The O-linked Glycosyltransferases GalNAcT2, Core1 and ST3Gal1 can be collectively used to add a core 1 structure including a terminal sialic acid or sialic acid-PEG onto a serine or threonine residues of selected proteins including therapeutic proteins. The expression of these enzymes in *E. coli* and recovery of active enzymes from refolding

inclusion bodies is useful for developing a cost effective scaleable process. Here we describe co-refolding two of these enzymes (MBP-GalNAcT2 and MBP-ST3Gal1) from *E. coli* inclusion bodies. The advantages of co-refolding include decreased reagent use and increased refolding efficiency.

[0390] Two strains of JM109 were selected for kanamycin resistance and determined to be carrying the appropriate expression plasmids that accumulate inclusion bodies of MBP-GalNAcT2 and MBP-ST3Gal1 upon induction with IPTG.

[0391] After separate overnight inductions with 1mM IPTG, bacterial cells pellets were resuspended at a ratio of 1g wet cell pellet per 10 mL of 20mM Tris pH 8, 5mM EDTA and lysed by mechanical disruption with two passes through a microfluidizer. Insoluble material, *i.e.*, the inclusion bodies or IB's, was pelleted by centrifugation at 7000xG at 4°C in the Sorvall RC3 for 30 minutes, and the supernatant discarded. A typical wash cycle consisted of completely resuspending the pellet in the wash buffer, and repeating the centrifugation for 15 minutes. The pellet was washed once in an excess volume (at least 10 mL (up to 20) per g original cell pellet) of high salt buffer (wash I: 10 mM Tris pH 7.4, 1 M NaCl, 5 mM EDTA), once in detergent buffer (wash II: 25 mM Tris pH 8, 100 mM NaCl, 1% Triton X100, 1% Na-deoxycholate, 5 mM EDTA), and (to remove traces of detergent in addition to washing the IBs) three times in wash buffer (wash III: 10 mM Tris pH 8, 5 mM EDTA).

[0392] Washed IB's were resuspended in solubilization buffer (8M Urea, 50mM BisTris pH 6.5, 5mM EDTA, 10mM DTT) and adjusted to 4mg/ml protein concentration. The urea protein solution was clarified by centrifugation at 14000xG for 5 minutes. Refolding was performed as a rapid dilution to 0.1mg/ml in refold buffer (55mM Tris pH 8.2, 10.56 mM NaCl, 0.44mM KCl, 2.2 mM $MgCl_2$, 2.2 mM $CaCl_2$, 0.055% Peg3350, 550mM L-Arginine, 1mM GSH, 100mcM GSSG) and stirred for 16hrs at 4°C.

[0393] The experimental system was set up as follows: Refolding conditions were held constant while refolding either MBP-GalNAcT2 and MBP-ST3Gal1 separately or co-refolding in the same vessel. Following refold, the mixtures were desalted by low speed centrifugation using G50 Sephadex equilibrated in 50mM BisTris pH 6.5, 75mM NaCl, 0.05% Tween 80. Soluble material was recovered after centrifugation at top speed in a microfuge.

[0394] To determine the yield of soluble enzyme following the refolding, each reaction was subjected to SDS-PAGE analysis followed by staining with coomassie blue to visualize the

polypeptides (Figure 27A). Results show that co-refolding increases the yield of soluble enzymes (lanes for co-refold vs separate).

[0395] In the second experiment, 6μg of the therapeutic protein IFα-2b was subjected to a one pot three enzyme pegylation reaction. The concentration of enzymes used for this reaction are 1.4ug MBP-GalNAcT2, 0.5mu BV Core1, and 1.4ug MBP-ST3Gal1, the sugars used are 1.2ug UDP-GalNac and 1.2ug UDP-Gal and 125ug 20K-Peg-CMP-NAN in a 20ul reaction for either 0hrs or 16hrs using the following reaction buffer (50mM MES pH 6.2, 150mM NaCl, 10mM MnCl2). The extent of pegylation of IFα-2b in half of the reaction was visualized by coomassie blue staining following SDS-PAGE (Fig. 27B). The experiment was designed to directly compare the one pot pegylation using either the combination of individually refolded enzymes to the co-refolded enzymes holding all other components and concentrations constant. Results demonstrate an increase in pegylated product in the reaction where the enzymes were co-refolded compared to being refolded separately (lane 5 vs 7).

Example 12: Enhanced expression of MBP-SiaA fusion protein

[0396] The SiaA gene from non-typeable *Haemophilus influenzae* was codon optimized for expression in *E. coli*. The gene was digested with NdeI and EcoRI, agarose gel purified and ligated into NdeI/ EcoRI digested pcWin2, prior to transformation into *E.coli* strain JM109 or W3110. Plasmid DNA was isolated from the recombinants and screened with NdeI and EcoRI. One JM109 and one W3110 colony each containing the correct construct was seeded into 2ml of animal free LB containing 10 μg/ml kanamycin sulfate and grown for 6h at 37°C at 250RPM of agitation. A 100 μl aliquot was removed, centrifuged 2 minutes at 10,000 x g and the supernatant discarded. This pellet was frozen at -20°C and represents the uninduced cell. IPTG was added to the remainder of the culture at a final concentration of 1 mM IPTG and incubated for 2h at 37°C and 250 RPM of agitation. A 100 μl aliquot was removed and processed as described (represents induced cell).

[0397] The restriction sites at the ends of the SiaA gene were changed to 5'BamHI and the 3' end maintained as EcoRI using PCR. The PCR product was digested with BamHI and EcoRI, purified by agarose gel electrophoresis and ligated into BamHI/EcoRI digested pcWin2MBP. JM109 cells transformed with this ligation reaction were cultured and plasmid DNA was isolated. The plasmid DNA was screened using BamHI and EcoRI. Three colonies containing the correct structure were seeded into 2ml of animal free LB containing 10 μg/ml kanamycin sulfate and grown for 6h at 37°C at 250RPM of agitation. A 100 μl

aliquot was removed, centrifuged 2 minutes at 10,000 x g and the supernatant discarded. This pellet was frozen at -20oC and represents the uninduced cell. IPTG was added to the remainder of the culture at a final concentration of 1 mM IPTG and incubated for 2h at 37°C and 250 RPM of agitation. A 100 µl aliquot was removed and processed as described (represents induced cell).

[0398] Each 100 µl aliquot was boiled 5 minutes in 100 µl of SDS-PAGE sample buffer containing 50 mM DTT. The samples were loaded on 4-20% acrylamide Tris-glycine gel (Invitrogen) and electrophoresed for ~2h, stained with Invitrogen Simply Blue Safestain, destained with water and scanned. Figure X shows undetectable levels of expression of native SiaA upon induction, whereas Figure Y shows high levels of expression of SiaA fused to MBP upon IPTG induction. This result demonstrates that MBP supplied from the pcWin2MBP vector drives high level protein expression.

[0399] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.